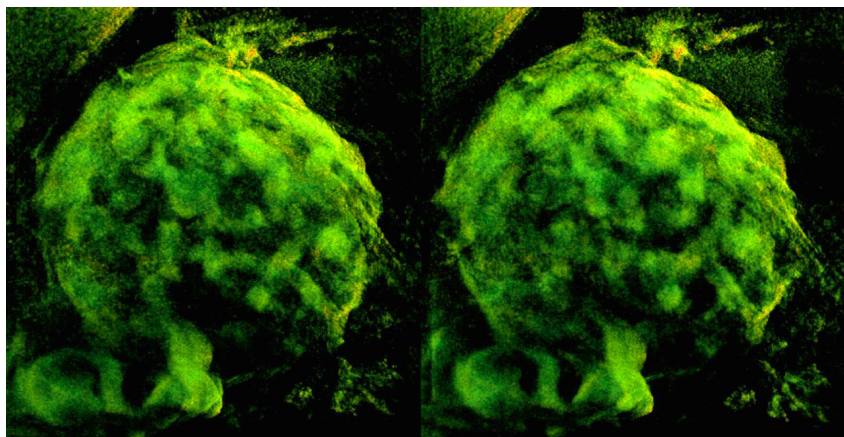


# THEOREMS AND ALGORITHMS FOR MULTIPLE VIEW GEOMETRY WITH APPLICATIONS TO ELECTRON TOMOGRAPHY

Sami Brandt



TEKNILLINEN KORKEAKOULU  
TEKNISKA HÖGSKOLAN  
HELSINKI UNIVERSITY OF TECHNOLOGY  
TECHNISCHE UNIVERSITÄT HELSINKI  
UNIVERSITE DE TECHNOLOGIE D'HELSINKI



# THEOREMS AND ALGORITHMS FOR MULTIPLE VIEW GEOMETRY WITH APPLICATIONS TO ELECTRON TOMOGRAPHY

Sami Brandt

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Electrical and Communications Engineering, for public examination and debate in Auditorium T2 at Helsinki University of Technology (Espoo, Finland) on the 18th, October, 2002, at 12 o'clock noon.

Helsinki University of Technology  
Department of Electrical and Communications Engineering  
Laboratory of Computational Engineering

Teknillinen korkeakoulu  
Sähkö- ja tietoliikennetekniikan osasto  
Laskennallisen tekniikan laboratorio

Helsinki University of Technology  
Laboratory of Computational Engineering  
P.O. Box 9203  
FIN-02015 HUT  
FINLAND  
Tel. +358 9 451 4826  
Fax. +358 9 451 4830  
<http://www.lce.hut.fi>

Available from <http://lib.hut.fi/Diss/2002/isbn9512261375/>

E-mail: [Sami.Brandt@hut.fi](mailto:Sami.Brandt@hut.fi)

© Sami Brandt

ISBN 951-22-6136-7 (printed)  
ISBN 951-22-6137-5 (PDF)  
ISSN 1455-0474  
Otamedia Oy  
Espoo, Finland 2002



# Abstract

The thesis considers both theory and algorithms for geometric computer vision. The framework of the work is built around the application of autonomous transmission electron microscope image registration.

The theoretical part of the thesis first develops a consistent robust estimator that is evaluated in estimating two view geometry with both affine and projective camera models. The uncertainty of the fundamental matrix is similarly estimated robustly, and the previous observation whether the covariance matrix of the fundamental matrix contains disparity information of the scene is explained and its utilization in matching is discussed. For point tracking purposes, a reliable wavelet-based matching technique and two EM algorithms for the maximum likelihood affine reconstruction under missing data are proposed. The thesis additionally discusses identification of degeneracy as well as affine bundle adjustment.

The application part of the thesis considers transmission electron microscope image registration, first with fiducial gold markers and thereafter without markers. Both methods utilize the techniques proposed in the theoretical part of the thesis and, in addition, a graph matching method is proposed for matching gold markers. Conversely, alignment without markers is disposed by tracking interest points of the intensity surface of the images. At the present level of development, the former method is more accurate but the latter is appropriate for situations where fiducial markers cannot be used.

Perhaps the most significant result of the thesis is the proposed robust estimator because of consistence proof and its many application areas, which are not limited to the computer vision field. The other algorithms could be found useful in multiple view applications in computer vision that have to deal with uncertainty, matching, tracking, and reconstruction. From the viewpoint of image registration, the thesis further achieved its aims since two accurate image alignment methods are suggested for obtaining the most exact reconstructions in electron tomography.

**Keywords:** robust regression, robust estimation, statistical modeling, epipolar geometry, fundamental matrix, uncertainty, image matching, affine reconstruction, affine triangulation, degeneracy, bundle adjustment, image registration, image alignment, computer vision, electron tomography.



# Contents

<b>Abstract</b>	<b>5</b>
<b>List of Abbreviations</b>	<b>11</b>
<b>List of Symbols</b>	<b>13</b>
<b>Preface</b>	<b>17</b>
<b>1 Introduction</b>	<b>19</b>
1.1 Background . . . . .	19
1.2 Research Problem . . . . .	20
1.3 Overview and Aims of the Thesis . . . . .	20
<b>2 Maximum Likelihood Robust Regression</b>	<b>23</b>
2.1 Background of the Regression Problem . . . . .	24
2.2 Outlier Free Regression Model . . . . .	26
2.3 Maximum Likelihood Robust Estimator . . . . .	28
2.4 Necessary Conditions . . . . .	34
2.5 EM Solution . . . . .	35
2.6 Unknown Residual Models . . . . .	37
2.7 Implementation Issues . . . . .	38
2.8 Summary and Discussion . . . . .	41
<b>3 Two View Geometry Estimation</b>	<b>43</b>
3.1 Introduction . . . . .	44
3.2 Affine F-Matrix Estimation . . . . .	46

---

3.3	Uncertainty of the Affine F-matrix . . . . .	49
3.4	Affine Experiments . . . . .	53
3.4.1	Synthetic Data . . . . .	53
3.4.2	Real Data . . . . .	56
3.5	Projective Experiments . . . . .	62
3.6	Summary and Discussion . . . . .	65
<b>4</b>	<b>Uncertainty of the F-Matrix and Disparity of the Scene</b>	<b>67</b>
4.1	Introduction . . . . .	67
4.2	Special Points and Lines . . . . .	69
4.3	How to Fix the Origin of the Image . . . . .	76
4.4	Epipolar Uncertainty and Matching . . . . .	80
4.5	Summary and Discussion . . . . .	86
<b>5</b>	<b>Multi-Resolution Image Matching</b>	<b>89</b>
5.1	Introduction . . . . .	89
5.2	Wavelet-Based Matching . . . . .	90
5.3	Experiments . . . . .	92
5.4	Summary . . . . .	95
<b>6</b>	<b>Affine Reconstruction and Missing Data</b>	<b>97</b>
6.1	Introduction . . . . .	97
6.2	Problem Definition . . . . .	99
6.3	Closed-Form Solution for the Translation Parameters . . . . .	100
6.4	Closed-Form Solution for the 3D Coordinates . . . . .	101
6.5	Application Examples . . . . .	104
6.5.1	Affine Triangulation . . . . .	104
6.5.2	Affine Reconstruction . . . . .	105
6.5.3	Degeneracy Identification . . . . .	108
6.5.4	Image Registration in Electron Tomography . . . . .	110
6.6	Experiments . . . . .	110
6.7	Note on Algorithmic Complexity . . . . .	114

---

6.8	Summary . . . . .	115
<b>7</b>	<b>TEM Image Alignment with Fiducial Markers</b>	<b>117</b>
7.1	Introduction . . . . .	118
7.2	Fiducial Marker Localization . . . . .	119
7.3	Marker Matching and Tracking . . . . .	120
7.3.1	Finding Candidates . . . . .	122
7.3.2	Graph Matching . . . . .	123
7.4	Parameter Optimization . . . . .	125
7.5	Experiments . . . . .	126
7.6	Further Development . . . . .	128
7.7	Summary and Discussion . . . . .	131
<b>8</b>	<b>TEM Image Acquisition Without Markers</b>	<b>133</b>
8.1	Introduction . . . . .	133
8.2	Methods . . . . .	134
8.3	Experiments . . . . .	135
8.4	Summary and Discussion . . . . .	143
<b>9</b>	<b>Conclusions</b>	<b>147</b>
	<b>Bibliography</b>	<b>149</b>



# List of Abbreviations

AIC	Akaike information criterion
a.s.	Almost surely
BA	Bilinear algorithm
BIC	Bayesian information criterion
CEM	Classification expectation maximization
cpd	Critical-point-dried
CMU	Carnegie Mellon University
EM	Expectation maximization
ET	Electron tomography
F	Fundamental (matrix)
GETA	Graduate School in Electronics, Telecommunications, and Automation
HUT	Helsinki University of Technology
i.i.d.	Independent and identically distributed
IF	Iterative factorization
INRIA	Institut national de recherche en informatique et en automatique
LMedS	Least median of squares
LS	Least squares
M	Maximum-likelihood-type (estimator)
MEM	Maximum entropy method
MDL	Minimum description length
ML	Maximum likelihood
MLE	Maximum likelihood estimate/estimator
MLESAC	Maximum likelihood sample consensus
MLRE	Maximum likelihood robust estimate/estimator
MSAC	M-estimator sample consensus
OR	Orthogonal regression
RANSAC	Random sample consensus
RMS	Root mean square
SEM	Stochastic expectation maximization
SVD	Singular value decomposition
SRM	Structural risk minimization
TEM	Transmission electron microscope
u.a.s.	Uniformly and almost surely
u.s.	Uniformly and surely

VASC	Vision and Autonomous Systems Center
WLS	Weighted least squares



# List of Symbols

$\hat{\cdot}$	Estimate
$\cdot^*$	Estimator
$\ \cdot\ _1$	Gerschgorin norm or the 1-norm
$\ \cdot\ _2$	Euclidean distance or the 2-norm
$\ \cdot\ _\infty$	Chebyshev or the maximum norm
$\hat{=}$	Correspondence between two symbols
$\cdot$	Definition
$\leftrightarrow$	Correspondence between two points
$\mathbf{0}$	Null vector
$A$	Area
$\mathbf{A}$	Coefficient matrix
$\alpha$	Rotation angle
$\beta$	Tilt angle
$C$	Total cost in least squares
$\cdot$	Cost function
$\mathbf{C}$	Coefficient matrix for a conic; Covariance matrix
$d$	Distance
$\mathbf{d}$	Unit direction vector
$\mathcal{D}$	Data set
$\delta$	Precision parameter
$\delta_{ij}$	Kronecker delta product
$\Delta\mathbf{F}$	Zhang's difference between two fundamental matrices
$g$	Model defining functional
$\gamma$	A priori probability
$\mathbf{e}$	Canonical basis vector
$E$	Expectation operator
$E_0$	Expectation with respect to the relevant distribution
$\epsilon$	Residual realization
$\epsilon_{\text{est}}$	RMS estimation error
$\mathbf{f}$	Parameter vector of the affine fundamental matrix
$\mathbf{F}$	Fundamental matrix
$\mathcal{F}$	Set of false measurements

---

$\mathbf{I}$	Identity matrix
$J$	Jacobian operator
$\kappa$	Condition number
$\text{Ker}$	Kernel operator
$\mathbf{l}$	Line
$\mathbf{l}_\infty$	Line at infinity
$l$	Length of the perimeter
$L$	Likelihood function
$\tilde{\cdot}$	Negative log-likelihood function
$\tilde{\cdot}_{\text{AIC}}$	Negative log-likelihood per data point
$\tilde{\cdot}_{\text{MDL}}$	Penalized negative log-likelihood in AIC
$\lambda$	Penalized negative log-likelihood in MDL
$\lambda$	Lagrange multiplier
$M$	Total number of feature points
$\mathbf{m}$	Point in an image
$\mathbf{M}$	Inhomogeneous projection matrix
$\mu$	Mean
$n$	Number of measurements
$\mathbf{n}$	Unit normal vector
$N$	Total number of images
$N_i$	Number of independent parameters
$N_t$	Total number of independent measurements
$O$	Upper bound for complexity
$\omega$	Model parameter
$\Omega$	Model parameter space
$\omega'$	Set of model and residual parameters
$p$	Probability density
$P$	Probability distribution
$\mathbf{P}$	Projection matrix
$\cdot$	Projective space
$\pi$	Projective plane
$\psi$	Residual parameter
$\Psi$	Residual parameter space
$\mathbf{q}$	Normalizing transform
$\cdot$	Real space
$\mathcal{R}$	Set of correct measurements
$r$	Signed distance
$\rho$	Correlation coefficient
$s$	Scale parameter
$\sigma$	Standard deviation
$\mathbf{t}$	Translation vector
$\mathbf{T}$	Homogeneous translation matrix
$\theta$	Direction angle

---

$\mathbf{U}$	Matrix of sorted unit eigenvectors $\mathbf{u}_1$ , $\mathbf{u}_2$ , and $\mathbf{u}_3$
$\text{Var}$	Variance operator
$\mathbf{W}$	Measurement matrix
$x, X$	Residual realization and the corresponding random variable (Chapter 2)
$x$	First image coordinate variable (except in Chapter 2)
$\mathbf{x}$	3D world coordinates of a point
$\chi$	Expected complete data log-likelihood
$y$	Second image coordinate variable
$\zeta$	Circularity



# Preface

This thesis is the result of my research at the Laboratory of Computational Engineering at Helsinki University of Technology during the years 1999–2002.

I got my first contact with geometry and computer vision when I was an undergraduate student, and started to work as a summer trainee at the Laboratory of Computational Engineering in 1997. Then, I carried out a special research study on camera calibration, and remember myself being full of enthusiasm for the subject. Yet, after two of years of searching for my interests and finishing my master’s thesis at the Laboratory of Computer and Information Science, I was offered the opportunity to carry out post-graduate research on 3D computer vision. I knew I had finally found what I was looking for and so I returned to the Laboratory of Computational Engineering in September 1999.

I am obliged to Prof. Kimmo Kaski for providing me with the opportunity to follow my research interests in computer vision. I want to express my deepest gratitude to Dr. Jukka Heikkonen, my research supervisor, for the many fruitful discussions, and helping me with my work in so many ways. While he has been busy with his own work throughout, he has always had time for a chat. It has been a pleasure to work with him. This work would not have been possible without Dr. Peter Engelhardt, who originally introduced us to the alignment problem of electron microscope images, and I also wish to thank him for that.

I am grateful to Mr. Juho Kannala and Mr. Tuomas Pansar, who share similar research interests to me. Their comments on my work and our numerous conversations have been invaluable. My sincerest thanks to Prof. Jorma Rissanen for valuable talks and comments concerning parts of the thesis. I would also like to thank Lecturer Simo K. Kivelä for our discussion and, above all, his life work of teaching mathematics; his enthusiasm and exactness initially evoked my fascination to linear algebra. Thanks to Mr. Markus Varsta and Dr. Aki Vehtari for their comments on part of the work, and Ms. Eeva Lampinen for taking care of the practicalities.

My special appreciation goes to my wife Tellervo for her love and understanding, for reading and commenting my texts, and sharing her expertise in mathematics. I also owe a great debt of gratitude to my mother and father for their encouragement during my studies.

I greatly acknowledge the funding by the Graduate School in Electronics, Telecommunications, and Automation (GETA); Academy of Finland, Research Centre for Computational Science and Engineering, project no. 44897 (Finnish Centre of Excellence Programme 2000–2005); and Tekniikan edistämissäätiö.

Espoo, 7th May 2002

Sami Brandt

# Chapter 1

## Introduction

We start this chapter by describing the background of the research area. The research problem is introduced in Section 1.2 and the aims of the thesis are set out in Section 1.3.

### 1.1 Background

Computer vision has emerged as an intensive discipline during the last four decades. Making a computer see is a challenging problem that is connected to the physics and mathematics of image formation, computer science, and, through biological vision systems, to cognition and neurosciences. Its ultimate goal, however, is not directly to model biological vision since, at least to present knowledge, how the biological vision system works is still unknown. Nevertheless, computer vision research has already achieved both practical and theoretical success.

The oldest branch of mathematics, geometry, provides laws relating to how multiple images of a scene are related. Any vision system is fundamentally related to geometry in this respect, and therefore it is natural that one field of computer vision is based on a geometric viewpoint. Projective geometry in particular, developed by mathematicians in the 19<sup>th</sup> century, accommodates elegantly the projections of conventional cameras in many respects. Perhaps the most important impetus for the recent two decades of intensive research into geometric computer vision was given by Longuet-Higgins in 1981, when he published his article of essential matrix in *Nature*.

New applications of geometric computer vision appear continually as only our imagination and ignorance of relevant practical problems slow down development. In terms of specific applications, so far there has been research into creating image mosaics, image synthesis, camera self-calibration, and autonomous or drive-assisted vehicles. In the future, augmented reality and wireless applications

should provide new possibilities. Even cellular phones have built-in cameras, and it remains to be seen what one will be able to do with them.

## 1.2 Research Problem

The research problem relating to this work is two-fold. From the direct scientific point of view, the purpose is to contribute new scientific results to the field of geometric computer vision. A convenient way achieving this objective is to consider an application where one has to use present knowledge and develop theory and algorithms when necessary. In the present study, the application is the registration problem of electron microscope images. Studying this application can also be seen as a fruitful scientific goal since, as far as a structural biologist is concerned, better registration methods imply better reconstructions and abilities to study and understand the structure of biological objects. Briefly described, the application problem is the following.

The purpose of electron tomography is to study small scale biological objects from their transmission electron microscope (TEM) images. In order to better study the spatial relationships and the structure of the objects, a researcher prefers a 3D reconstruction to a 2D image series. Since the TEM images are transparent, the reconstruction problem can be solved via standard methods of tomography if the motion of the object between the images is known. Our research problem culminates in solving the motion from the image series automatically since the motion is initially unknown because of the nature of the imaging process.

From the computer vision point of view, the application problem has several interesting sub problems. For instance, many geometric problems in computer vision involve robust regression since the measurements usually contain outliers. Important questions include whether present methods of robust estimation are sound enough, and more importantly, how good they can theoretically be. Since in practice we must deal with noisy measurements, the estimated models involve a degree of uncertainty. Other questions relate to the information that the uncertainty contains, and how it could be utilized.

## 1.3 Overview and Aims of the Thesis

The primary goals of this thesis are to obtain new scientific knowledge for the discipline of computer vision and to solve the automatic registration problem of TEM tilt series. The thesis is organized so that we consider the sub problems related to the application individually, and finally collect the methods, complete the missing parts, and evaluate the resulting TEM image registration methods in



the application domain. The detailed partial goals of each chapter are described below.

Since robust estimation is crucial in image geometry estimation, Chapter 2 considers problems of previously proposed methods of robust regression, and its aim is to propose a robust estimator that is a consistent estimator of true parameter values regardless of problematical outliers in the observations. The chapter forms the most theoretical part of the thesis, and the goal is not only to describe the properties of such an estimator but also to discuss and give an algorithm for its implementation.

Chapter 3 covers two view geometry estimation, where the objective is to evaluate the proposed robust estimator in practice. The estimation of the two view tensor, the fundamental matrix, has attracted much attention in the last two decades. It has been claimed that its estimation problem has already been solved, though the state-of-the-art methods rely on the previous methods of robust estimation that continue to entail certain problems (see Chapter 2). A further objective of the chapter is to characterize the uncertainty of the fundamental matrix in terms of its covariance estimation, without permitting the outliers to violate the noise distribution assumptions.

The aim of Chapter 4 is to test the earlier hypothesis regarding whether the covariance matrix of the fundamental matrix reveals something about the disparity of the scene. An important question is how this information could be utilized from the matching point of view. An additional goal is therefore to show that, in addition to the point–line relation of the epipolar geometry, there is a slightly stronger relation if the covariance information of the fundamental matrix is used. The analysis will be chiefly theoretical but we will provide some illustrative examples with real images.

In Chapter 5, the objective is to obtain reliable point matches in two images by using the multi-resolution information of the intensity surface, and utilizing the epipolar geometry and its uncertainty. The matches should be practically free from mismatches when natural images are used. The approach is validated experimentally.

The main target of Chapter 6 is to derive a closed-form solution for the structure and translation parameters for a motion sequence imaged by an affine camera. We use the result as an initial estimate in bundle adjustment where structure and constrained motion parameters are optimized simultaneously. Secondary aims of the chapter include a consideration of other by-products of the result. We consider the problem of maximum likelihood affine reconstruction with missing data and its solution with iterative algorithms. Additionally studied is whether the result can discover degenerate motion or structure configurations.

The aim of Chapter 7 is to summarize the proposed techniques and solve the

TEM image alignment problem when fiducial gold markers can be used. The secondary goal of the chapter is to determine how gold markers can be localized and tracked from the tilt-series. A further objective is to evaluate the registration method with real tilt-series and assess its applicability and possibilities for future development.

In Chapter 8, we consider the more difficult case of TEM image alignment, i.e., when fiducial markers cannot be used. The techniques of that chapter are proposed earlier in the thesis, and the main goal is therefore to evaluate experimentally how the combination of these techniques performs in the alignment problem. Future directions in application development are also outlined.

In Chapter 9, we discuss the principal results of the thesis.

## Chapter 2

# Maximum Likelihood Robust Regression<sup>1</sup>

*Everyone believes in the [normal] law of errors, experimenters because they think it is a mathematical theorem, and mathematicians because they think it is an experimental fact.*

Henri Poincare

Robust model estimation is a central problem in many research areas, including computer vision. The unsolved difficulty has been in dealing with *outliers*, rogue observations that are not explicable by the underlying model. Even though a vast amount of research effort has been dedicated to the problem it is not known how to deal with contaminants in the regression data in an *optimal* way. By an optimal robust estimate, we mean the *same* estimate that would be achieved *without* contaminants in the data. In this chapter, we will propose an estimator that is asymptotically optimal. Principally, we will show proofs and conditions for optimality, consistency and unbiasedness for the proposed estimator.

The proposed framework can be divided into two branches. The easier case occurs if the source of the contaminants is known and it is possible to give a parametric model for their residual distribution. Then, the optimal estimate can be computed with a simple expectation maximization (EM) algorithm. The case is more difficult when the form of the contaminant residual distribution is unknown. In such a case, the total residual density of the model must be estimated using an appropriate set of basis functions and the complexity of the model must be controlled. In effect, optimal complexity control is provided by the structural risk minimization principle (SRM) (Vapnik 2000) and minimum description length (MDL) (Rissanen 1989), but other techniques exist.

---

<sup>1</sup>The content of this chapter has been published in shorter form in Brandt (2002b).

This chapter forms the most mathematical part of this thesis. It introduces the principle known as maximum likelihood robust regression which is applied and further evaluated in the epipolar geometry estimation in Chapter 3. Section 2.1 provides the background to the robust regression problem, while the formal definitions, theorems and proofs are given in Sections 2.2 and 2.3. Implementation of the principle is then discussed by first reporting the necessary conditions for the solution (Section 2.4), proposing a simple iterative algorithm to the case where the form of the outlier distribution is known (Section 2.5), and the most general case of unknown residual models is finally discussed in Section 2.6. Other implementation issues are considered in Section 2.7; a summary and discussion follow in Section 2.8.

## 2.1 Background of the Regression Problem

The least squares method (LS) is a widely used technique in data analysis because of its simplicity and optimality under Gaussian noise in data. Whatever the case may be, the problem can be defined in many ways. The simplest formulation is known as 'ordinary least squares', where independent and identically distributed (i.i.d.) additive Gaussian noise on the dependent variable is assumed. A more general formulation is known as 'orthogonal regression' (OR), 'principal component regression', or 'total least squares regression', where it is assumed that *all* the variables are corrupted by the i.i.d. Gaussian noise. Another taxonomy considers whether the underlying model is linear or non-linear. Linear least squares can be seen as hyper plane fitting whereas non-linear least squares seek for parameters for a functional relationship between the variables where the form of the functional relationship or *model* is known.

The least squares procedure is inadequate if the observations are contaminated other than by i.i.d. Gaussian noise. However, if the noise distribution is normal but with varying covariance matrix, there is a straightforward modification, the weighted least squares (WLS), where the square distance cost is replaced by the squared Mahalanobis distance. On the other hand, minimizing the Chebyshev or maximum norm  $\|\cdot\|_\infty$  is optimal for uniformly distributed residual, in the maximum likelihood sense, and is therefore even more vulnerable to outliers than the traditional least squares. Conversely, the Gerschgorin or the 1-norm  $\|\cdot\|_1$  is optimal for Laplace or double exponential distribution, and is therefore less sensitive to outliers than least squares.

The problem of *outliers* has been problematic in the identification of functional relationships in the extensive range of scientific research. From a statistician's perspective, outliers are understood as the data that are in gross disagreement with the postulated model. The problem is well known as *outlier rejection*, *robust estimation*, or *robust regression*, and it has been widely researched. Indeed,

numerous books have been written on the subject alone; a quick search in Amazon.com bookstore reveals that at least 15 books have been published on the subject since 1980.

The methods proposed for resolving the outlier problem can be divided into two classes (Rousseeuw and Leroy 1987). The basic idea of *regression diagnostics* involves first identifying and removing outliers according to some rule of outlier rejection, and then reanalyzing the remaining data. This can also be an iterative process, and is therefore occasionally known as the “build and criticize” approach. The other class of methods is called *robust statistics*, where the goal is to devise estimators that are tolerant of large residual values. After computing the robust estimate, one is supposed to identify the outliers by examining the residual values.

The classical method in regression diagnostic involves first computing a least-squares fit, then removing the points whose residuals exceed a pre-defined threshold, and iterating until the outliers are removed. This may work well if the number of outliers is small but even a single outlier far from data centroid can be fatal since it may act as a *leverage point* (Rousseeuw and Leroy 1987) and have a small residual value in the fit with respect to the others. A more effective method is to use influence measures to pinpoint potential outliers. The difference is that, while classical methods examine the current residual in rejecting outliers, these methods search for the solution that gives the best results in the next iteration. A good example is the Shapiro and Brady’s (1995) method in the OR framework. However, if the number of contaminants is large, robust statistics has been found to work better.

Robust statistics has the same goal as regression diagnostics, but the procedure goes in reverse order since after the good *robust* fit, the outliers are to be identified if necessary. One of the best known methods is ‘least median of squares’ (LMedS) (Rousseeuw 1984), which geometrically involves searching for the narrowest strip that contains half of the residuals, and it is normally followed by an M-estimator (Rousseeuw and Leroy 1987). M-estimators (maximum likelihood type) (Huber 1981) are another group of methods, where the square function in LS is replaced with another symmetric, positive definite function that obtains unique minimum at zero. Furthermore, among the computer vision community, certain popular methods are based on RANSAC (random sample consensus) (Fischler and Bolles 1981, Torr and Zisserman 2000), where minimal configurations are randomly sampled and one selects the one that maximizes the number or some other quantity of residuals, often denoted as *inliers*, that fall inside some predefined threshold.

The known procedures, however, suffer from the following problems:

1. In robust statistics the residual form is fixed in advance. Therefore, it is not adaptive in any way to the contaminating residuals that however may be in some sense independent from the relevant ones.

2. Methods in robust statistics assume symmetric residual distribution. There is no reason for this in general since the contaminants may be entirely asymmetrically distributed, which causes the estimates to be biased.
3. Even if the original “good” residuals alone are normally distributed, the residual after the outlier removal is not, since there will be false rejections and false retentions.
4. The classification of residual values to good and false is an ill-posed problem since the false observations can also have small residual values. This is because the distribution of good and contaminating observations residuals overlap in general.

To cope with the above problems, we must define an outlier in a slightly different way to how they are usually considered. An exact outlier definition is a little problematic. Rousseeuw and Leroy (1987) demonstrated this by *leverage points* that might be considered outliers but, on the contrary, whose residuals are small implying that they are not *regression outliers*. This could happen, for instance, when a line is fitted to points and one observation is far away from the other points but still, by chance, close to the true line. Consequently, it is not possible to find such outliers simply by examining the residual values.

In this thesis, outliers are understood as the *residuals of false observations*. False observations are the contaminating data that are generated by another unintended stochastic process. Accordingly, this definition allows that the outliers generally have a distribution that *overlaps* the residual of the good observations and as an outlier does *not* imply a large residual value. In addition, there is no reason to assume that the *outlier distribution* would be symmetric with respect to the correct distribution, or that the outliers would have the same mean as the contaminant free residual.

From this view point, we will introduce our robust estimator that solves the problems described above. Above all, the estimator will be proven consistent, i.e., the model parameter estimates approach the true values *regardless of outliers in the data* as the number of measurements approaches infinity. We start by formulating the regression problem with no outliers. We will define a model in a general way so that the setting covers both linear and nonlinear regression problems.

## 2.2 Outlier Free Regression Model

First, we need a mathematical description for a model. We obtain this by associating the model with a functional (not necessarily linear) as follows.

**Definition 2.2.1** Let  $z \in Z$  be an observation and  $\omega \in \Omega$  the set of model parameters where  $Z$  and  $\Omega$  are metric spaces. The model is defined as the functional  $g : Z \times \Omega \longrightarrow \mathbb{R}$  for which holds: the model explains the observation  $z$  perfectly with parameters  $\omega \Leftrightarrow g(z, \omega) = 0$ .

If  $Z = \mathbb{R}^n$  and given the parameters  $\omega$ , a model can be understood as a hyper surface in  $\mathbb{R}^n$ . The model residual is defined as follows.

**Definition 2.2.2** Let the null space of the model  $g$  with respect to the first argument be  $\text{Ker}\{g(\omega)\}$  and  $d(z, y)$  a distance function in the observation space  $Z$ . The residual corresponding to the observation  $z \in Z$  is defined as the signed distance between the observation  $z$  and the null space, i.e.,

$$s \inf_{y \in \text{Ker}\{g(\omega)\}} d(z, y), \quad (2.1)$$

where  $s = \text{sign } g(z, \omega)$ .

The distribution of the residual depends on the selection of distance function  $d(z, y)$ , sign of  $g(z, \omega)$ , as well as the type of noise in  $z$ . A convenient selection for the distance function (and the model defining mapping  $g$  in terms of its sign) would be such that the residual obeyed mean zero normal distribution but at least its parametric form should be known. The following examples illustrate this.

**Example 2.2.3** Consider a line in  $\mathbb{R}^2$ . Let the line parameter vector be  $\mathbf{l} = (l_1 \ l_2 \ l_3)^T$  and points on the line (in projective form)  $\mathbf{m} = (x \ y \ 1)^T$ , i.e.  $\mathbf{l}^T \mathbf{m} = 0$ . We choose  $g(\mathbf{m}, \mathbf{l}) = \mathbf{l}^T \mathbf{m}$  and Euclidean distance for the metric. The residual for a point  $(x \ y)^T \in \mathbb{R}^2$  is the signed orthogonal distance from the line  $\mathbf{l}$ . Further, assume that the observations are corrupted by additive, independent and identically distributed (i.i.d.) Gaussian noise with covariance matrix  $\mathbf{C} = \sigma^2 \mathbf{I}$  where  $\mathbf{I}$  is an identity matrix in  $\mathbb{R}^2$ . Then, the residual is similarly normally distributed and the problem of finding  $\mathbf{l}$  by minimizing the sum of squared residuals is known as the 'orthogonal least squares problem'.

**Example 2.2.4** Assume that the observations  $\mathbf{z}_i \in \mathbb{R}^n$  are corrupted by additive, independent Gaussian noise with a general  $n \times n$  covariance matrix  $\mathbf{C}_i$  that is known but different for each observation. With an appropriate selection of  $g$ , the residual is independent, normally and identically distributed if we set the distance function  $d(\mathbf{z}_i, \mathbf{y})$  to be the Mahalanobis distance  $((\mathbf{z}_i - \mathbf{y})^T \mathbf{C}_i^{-1} (\mathbf{z}_i - \mathbf{y}))^{1/2}$ .

Now we are ready to define the regression problem.

**Definition 2.2.5 (Outlier Free Regression Problem)** Let  $g$  be a model and  $x_i, i = 1, \dots, n$  independent and identically distributed residuals defined as in (2.1) for the data set  $\mathcal{D}$ . Let  $\omega' = \{\omega, \psi\}$  be the set of both model and residual

density parameters<sup>2</sup> as  $\omega \in \Omega$ ,  $\psi \in \Psi$ . The problem is to find the parameter set  $\omega'$  that maximizes the likelihood function

$$L(\mathcal{D}|\omega') = \prod_{i=1}^n p(x_i|\omega') \quad (2.2)$$

or equivalently minimizes the negative log-likelihood function

$$(\mathcal{D}|\omega') = - \sum_{i=1}^n \log p(x_i|\omega') \quad (2.3)$$

where  $p(x|\omega')$  is the probability density function of  $x$ .

Hence, the outlier free regression problem searches for the maximum likelihood estimate (MLE) for the model  $g$ .

## 2.3 Maximum Likelihood Robust Estimator

When the data are contaminated by outliers, we will not be able to determine precisely which observations are good and false. However, as we will show, we are able to construct an estimator that is consistent with the *true* parameter values. We do this by constructing a consistent and unbiased estimator, given the residual models and contaminated data, for the expectation value of (2.3) normalized by the number of observations. In fact, the proposed estimator then also approaches the MLE, given the good observations since the MLE minimizes (2.3). Therefore, the estimator is named as the maximum likelihood robust estimator (MLRE). Formally, the problem is defined as follows:

**Definition 2.3.1 (Problem)** *Let  $\mathcal{D}$  be a set of  $n$  observations that are generated by a Bernoulli process with the constant parameter  $\gamma_{\mathcal{R}}$  from relevant and false observations.  $\mathcal{D}$  then consists of the corresponding subsets  $\mathcal{R}$  and  $\mathcal{F}$ , where  $\mathcal{R}$  is the set of correct measurements that follow the model  $g$  up to a certain additive measurement noise and  $\mathcal{F}$  is the set of false measurements. Given parameter set  $\omega \in \Omega$  for the model  $g$ , each datum in  $\mathcal{D}$  produces a residual that is characterized by the random variable  $X$  that follows the mixture distribution*

$$p(x|\psi, \omega) = \gamma_{\mathcal{R}}p(x|\mathcal{R}, \psi_{\mathcal{R}}, \omega) + (1 - \gamma_{\mathcal{R}})p(x|\mathcal{F}, \psi_{\mathcal{F}}, \omega), \quad (2.4)$$

where  $p(x|\mathcal{R}, \psi_{\mathcal{R}}, \omega)$  and  $p(x|\mathcal{F}, \psi_{\mathcal{F}}, \omega)$  are the residual distributions<sup>3</sup> for relevant and false measurements, and  $\psi = \{\gamma_{\mathcal{R}}, \psi_{\mathcal{R}}, \psi_{\mathcal{F}}\} \in \Psi$  is the set of residual parameters. The problem is to determine the estimator  $\omega^*$  for  $\omega$  given the data  $\mathcal{D}$  and

<sup>2</sup>For instance, if the residual noise model is a single mean-zero Gaussian, the residual parameter  $\psi$  consists of the residual deviation  $\sigma$ .

<sup>3</sup>A more correct notation for  $p(x|\mathcal{R}, \psi_{\mathcal{R}}, \omega)$  would be  $p(x|x \in \mathcal{R}, \psi_{\mathcal{R}}, \omega)$  but the former is used for brevity. The same applies to the forthcoming posterior probabilities.



density models  $p(x|\mathcal{R}, \psi_{\mathcal{R}}, \omega)$  and  $p(x|\mathcal{F}, \psi_{\mathcal{F}}, \omega)$  so that  $\omega^*$  is consistent, that is, it approaches  $\omega_0$  as  $n \rightarrow \infty$ , where  $\omega_0$  is the “true” parameter set with which the relevant data have been generated.

For simplicity, we assume from now on, unless otherwise stated, that the residual noise obeys i.i.d. mean-zero Gaussian distribution, in which case  $\psi_{\mathcal{R}} = \{\sigma_{\mathcal{R}}\}$ . The problem is mainly solved by proving the following theorem.

**Theorem 2.3.2** *Consider the estimation of parameters  $\omega$  for the model  $g$  from the data  $\mathcal{D}$ , where  $\mathcal{D}$  is a set of  $n$  observations that are generated by a Bernoulli process with constant parameter  $\gamma_{\mathcal{R}}$  from relevant and false observations. Let the residuals of the elements in  $\mathcal{D}$  follow the model (2.4), where the relevant and false densities are continuous in  $\omega' = \{\psi, \omega\} \in \Omega'$  for almost all  $x$ , and let  $\Omega' \subset \Psi \times \Omega$  be closed and bounded. Define a cost function to the data  $\mathcal{D}$  with the i.i.d. residuals  $x_1, \dots, x_n$  by*

$$(\mathcal{D}|\omega') = \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \sum_{i=1}^n P(\mathcal{R}|x_i, \omega') x_i^2 + \log \sqrt{2\pi} \sigma_{\mathcal{R}}, \quad (2.5)$$

where  $\sigma_{\mathcal{R}}^2$  is the variance of the relevant residuals and  $P(\mathcal{R}|x_i, \omega')$  is the posterior probability of a sample being relevant after examining its residual  $x_i$  given by the Bayes rule

$$P(\mathcal{R}|x_i, \omega') = \frac{\gamma_{\mathcal{R}} p(x_i|\mathcal{R}, \psi_{\mathcal{R}}, \omega)}{\gamma_{\mathcal{R}} p(x_i|\mathcal{R}, \psi_{\mathcal{R}}, \omega) + (1 - \gamma_{\mathcal{R}}) p(x_i|\mathcal{F}, \psi_{\mathcal{F}}, \omega)}. \quad (2.6)$$

Let the residual densities be such that  $P(\mathcal{R}|x_i, \omega')$  is continuous in  $x$  almost everywhere. Let  $\tilde{\ell}(\mathcal{R}|\psi_{\mathcal{R}}, \omega) = n_{\mathcal{R}}^{-1} \ell(\mathcal{R}|\psi_{\mathcal{R}}, \omega)$  be the negative log-likelihood function per data point of the relevant data, and assume that it converges uniformly and almost surely on  $\psi_{\mathcal{R}} \times \Omega$  to its expected value, assuming that the limit function exists and is continuous in  $\omega'$  for almost all  $x$ , where the expectation is taken with respect to the relevant distribution with density  $p(\mathcal{R}|\psi_{0,\mathcal{R}}, \omega_0)$ . Then,

$$(\mathcal{D}|\omega') \xrightarrow[n \rightarrow \infty]{a.s.} E_0\{\tilde{\ell}(\mathcal{R}|\psi_{\mathcal{R}}, \omega)\}, \text{ uniform on } \Omega', \quad (2.7)$$

that is, the cost converges uniformly and almost surely to the expected negative log-likelihood per data point evaluated at the relevant data.

*Proof.* The sum in (2.5) may be partitioned to intervals  $d_j = [x_j, x_j + \Delta x)$  of equal length, where  $\cup_{j \in \mathbb{Z}} d_j = \mathbb{R}$  and  $d_i \cap d_j = \emptyset$ ,  $i \neq j$ . By denoting  $x_i = x_j + x_i - x_j = x_j + \Delta x_i$ , where  $0 \leq \Delta x_i < \Delta x$ , and  $P(\mathcal{R}|x_i, \omega') = P(\mathcal{R}|x_j, \omega') +$

$P(\mathcal{R}|x_i, \omega') - P(\mathcal{R}|x_j, \omega') = P(\mathcal{R}|x_j, \omega') + \Delta P(\mathcal{R}|x_i, \omega')$ . We get

$$\begin{aligned}
(\mathcal{D}|\omega') &= \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \sum_j \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') x_i^2 + \log \sqrt{2\pi} \sigma_{\mathcal{R}} \\
&= \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \sum_j \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') (x_j^2 + 2x_j \Delta x_i + (\Delta x_i)^2) + \log \sqrt{2\pi} \sigma_{\mathcal{R}} \\
&= \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \sum_j (x_j^2 \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') + \\
&\quad + 2x_j \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') \Delta x_i + \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') (\Delta x_i)^2) + \log \sqrt{2\pi} \sigma_{\mathcal{R}} \\
&= \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \sum_j \left( x_j^2 \left( \sum_{i|x_i \in d_j} P(\mathcal{R}|x_j, \omega') + \sum_{i|x_i \in d_j} \Delta P(\mathcal{R}|x_i, \omega') \right) + \right. \\
&\quad \left. + 2x_j \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') \Delta x_i + \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') (\Delta x_i)^2 \right) + \log \sqrt{2\pi} \sigma_{\mathcal{R}}.
\end{aligned} \tag{2.8}$$

Let  $n_{j,\mathcal{R}}$  and  $n_{j,\mathcal{F}}$  be the number of relevant and false residuals on the interval  $d_j$ , respectively. On  $d_j$ , the relative number of correct residuals  $n_{j,\mathcal{R}}/(n_{j,\mathcal{R}} + n_{j,\mathcal{F}})$  approaches  $P(\mathcal{R}|d_j, \omega') = \int_{d_j} p(x|\mathcal{R}, \psi_{\mathcal{R}}, \omega) dx / \int_{d_j} p(x|\omega') dx$  almost surely according to the law of large numbers. The convergence is also uniform because  $\Omega'$  is closed and bounded, and  $P(\mathcal{R}|d_j, \omega')$  continuous in  $\omega'$  for almost all  $x$ . Thus, by first letting  $n \rightarrow \infty$  and finally  $\Delta x \rightarrow 0$  we get

$$\begin{aligned}
&\lim_{\Delta x \rightarrow 0} \lim_{n \rightarrow \infty} (\mathcal{D}|\omega') \\
&= \lim_{\Delta x \rightarrow 0} \lim_{n \rightarrow \infty} \frac{\overbrace{n_{\mathcal{R}}/n}^{\text{u.a.s.} \rightarrow \gamma_{\mathcal{R}}}}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \sum_j x_j^2 n_{j,\mathcal{R}} \underbrace{\frac{n_{j,\mathcal{R}} + n_{j,\mathcal{F}}}{n_{j,\mathcal{R}}}}_{\text{u.a.s.} \rightarrow 1/P(\mathcal{R}|d_j, \omega')} \underbrace{\frac{\sum_{i|x_i \in d_j} P(\mathcal{R}|x_j, \omega')}{n_{j,\mathcal{R}} + n_{j,\mathcal{F}}}}_{=P(\mathcal{R}|x_j, \omega')} + \\
&\quad + \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \left( \sum_j x_j^2 \sum_{i|x_i \in d_j} \Delta P(\mathcal{R}|x_i, \omega') + 2 \sum_j x_j \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') \Delta x_i + \right. \\
&\quad \left. + \sum_j \sum_{i|x_i \in d_j} P(\mathcal{R}|x_i, \omega') (\Delta x_i)^2 \right) + \log \sqrt{2\pi} \sigma_{\mathcal{R}}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{\Delta x \rightarrow 0} \frac{1}{2\sigma_{\mathcal{R}}^2} E_0 \left\{ x_j^2 \underbrace{\frac{P(\mathcal{R}|x_j, \omega')}{P(\mathcal{R}|d_j, \omega')}}_{\xrightarrow{\text{u.a.s.}} 1} | \mathcal{R} \right\} + \\
&\quad + \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}}} \left( E_0 \left\{ x_j^2 \underbrace{\Delta P(\mathcal{R}|x_i, \omega')}_{\xrightarrow{\text{u.a.s.}} 0} \right\} + 2E_0 \left\{ x_j P(\mathcal{R}|x_i, \omega') \underbrace{\Delta x_i}_{\xrightarrow{\text{u.s.}} 0} \right\} + \right. \\
&\quad \left. + E_0 \left\{ P(\mathcal{R}|x_i, \omega') (\underbrace{\Delta x_i}_{\xrightarrow{\text{u.s.}} 0})^2 \right\} \right) + \log \sqrt{2\pi} \sigma_{\mathcal{R}} \text{ u.a.s.}, \\
&= E_0 \{ \tilde{(\mathcal{R}|\psi_{\mathcal{R}}, \omega)} \} \text{ u.a.s.},
\end{aligned} \tag{2.9}$$

since  $\tilde{(\mathcal{R}|\psi_{\mathcal{R}}, \omega)} \xrightarrow[n_{\mathcal{R}} \rightarrow \infty]{\text{u.a.s.}} E_0 \{ \tilde{(\mathcal{R}|\psi_{\mathcal{R}}, \omega)} \}$  by assumption.

This theorem suggests that the maximum likelihood robust estimate could be computed by minimizing the cost function (2.5). This is indeed true since the estimate is consistent. To prove this formally, we yet need a slightly stronger condition than the minimum being unique.

**Definition 2.3.3** *Let the function  $f(x)$  have a unique minimum at  $x = x_0$ . The minimum is well separated if and only if for all  $\varepsilon > 0$*

$$\inf_{|x - x_0| \geq \varepsilon} f(x) > f(x_0). \tag{2.10}$$

**Theorem 2.3.4** *Let  $\hat{\omega}'_n$  be the estimate for  $\omega'$  that minimizes (2.5). Let  $(\mathcal{D}|\omega')$  converge uniformly and almost surely to the expected negative log-likelihood per data point of the relevant data whose minimum at  $\omega' = \omega'_0$  is unique and well separated. Then,*

$$\hat{\omega}'_n \xrightarrow[n \rightarrow \infty]{P} \omega'_0. \tag{2.11}$$

**Observation 2.3.5** *The theorem above states that by minimizing the proposed cost function the maximum likelihood estimate can be asymptotically achieved regardless of outliers in the data!*

*Proof.* The proof is similar to the proof of the weak consistence of the maximum likelihood estimator. Hence, let us denote the cost function with  $n$  observations by  $\ell_n(\omega')$  and the expected negative log-likelihood per data point by  $\ell(\omega')$ . Then,

$$\ell_n(\hat{\omega}'_n) \leq \ell_n(\omega'_0). \tag{2.12}$$

Since  $\ell_n(\omega'_0) \xrightarrow{P} \ell(\omega'_0)$ , there is a sequence of non-negative random variables  $\{Z_n\}$  with  $Z_n \xrightarrow{P} 0$  and  $|\ell_n(\omega'_0) - \ell(\omega'_0)| \leq Z_n$ . Thus, we have

$$|\ell(\hat{\omega}'_n) - \ell(\omega'_0)| \leq |\ell(\hat{\omega}'_n) - \ell_n(\omega'_0)| + Z_n \leq \sup_{\omega' \in \Omega'} |\ell(\omega') - \ell_n(\omega')| + Z_n, \tag{2.13}$$

where the right hand side converges to zero in probability. Because the minimum is unique and well separated,  $\forall \varepsilon > 0 \exists \delta > 0$  such that  $d(\hat{\omega}'_n, \omega'_0) \geq \varepsilon$  implies  $(\omega'_0) < (\hat{\omega}'_n) - \delta$ . Hence

$$P(d(\hat{\omega}'_n, \omega'_0) \geq \varepsilon) \leq P((\omega'_0) < (\hat{\omega}'_n) - \delta), \quad (2.14)$$

where the right hand side converges to zero.

Define  $\hat{\omega}'_n(\mathcal{D}|\omega') = \gamma_{\mathcal{R}} n \hat{\omega}'_n(\mathcal{D}|\omega')$ . Clearly, the same  $\omega$  that minimizes  $\hat{\omega}'_n(\mathcal{D}|\omega')$ , also minimizes  $\hat{\omega}'_n(\mathcal{D}|\omega')$ . Let us now replace the realizations with random variables in the functions. The unbiased nature of the estimator is characterized by the following theorem.

**Theorem 2.3.6** *The estimator  $\omega^*$  is unbiased in the sense that the expectation value of  $\hat{\omega}'_n(\mathcal{D}|\omega')$  is minimized with the same value  $\omega$  that minimizes the expectation value of  $\hat{\omega}'_n(\mathcal{R}|\psi_{\mathcal{R}}, \omega)$ , regardless of the number of measurements. Moreover,  $E\{\hat{\omega}'_n(\mathcal{D}|\psi, \omega)\} = E\{\hat{\omega}'_n(\mathcal{R}|\psi_{\mathcal{R}}, \omega)\} \forall n$ .*

*Proof.* Let  $N_{\mathcal{R}}$  be the random variable of the number of relevant measurements when the size of the data  $\mathcal{D}$  is  $n$ . Then, the log-likelihood estimator computed from the relevant data only is

$$\hat{\omega}'_n(\mathcal{R}|\psi_{\mathcal{R}}, \omega) = \frac{1}{2\sigma_{\mathcal{R}}^2} \sum_{i|X_i \in \mathcal{R}} X_i^2 + N_{\mathcal{R}} \log \sqrt{2\pi}\sigma_{\mathcal{R}}, \quad (2.15)$$

where the first term is a *random sum* over  $N_{\mathcal{R}}$  elements since the  $X_i$ :s are independent and identically distributed by definition. Hence,

$$E\{\hat{\omega}'_n(\mathcal{R}|\psi_{\mathcal{R}}, \omega)\} = \frac{1}{2\sigma_{\mathcal{R}}^2} E\{X^2|\mathcal{R}, \psi_{\mathcal{R}}, \omega\} E\{N_{\mathcal{R}}\} + E\{N_{\mathcal{R}}\} \log \sqrt{2\pi}\sigma_{\mathcal{R}}. \quad (2.16)$$

On the other hand, Bayes rule gives

$$P(\mathcal{R}|X, \psi, \omega) = \frac{\gamma_{\mathcal{R}} p(X|\mathcal{R}, \psi, \omega)}{\gamma_{\mathcal{R}} p(X|\mathcal{R}, \psi, \omega) + (1 - \gamma_{\mathcal{R}}) p(X|\mathcal{F}, \psi, \omega)} \quad (2.17)$$

$\Leftrightarrow$

$$p(X|\psi, \omega) = \frac{\gamma_{\mathcal{R}} p(X|\mathcal{R}, \psi, \omega)}{P(\mathcal{R}|X, \psi, \omega)}. \quad (2.18)$$

By taking the expectation value of  $Q^*(\mathcal{D}|\psi, \omega)$  and finally using (2.16), we get

$$\begin{aligned}
E\{Q^*(\mathcal{D}|\psi, \omega)\} &= \int_{-\infty}^{\infty} Q^*(\mathcal{D}|\psi, \omega) p(x|\psi, \omega) dx \\
&= \frac{1}{2\sigma_{\mathcal{R}}^2} \int_{-\infty}^{\infty} \sum_{i=1}^n P(\mathcal{R}|x_i, \psi, \omega) x_i^2 p(x_i|\psi, \omega) dx_i + \gamma_{\mathcal{R}} n \log \sqrt{2\pi} \sigma_{\mathcal{R}} \\
&= \frac{\gamma_{\mathcal{R}}}{2\sigma_{\mathcal{R}}^2} \sum_{i=1}^n \int_{-\infty}^{\infty} x_i^2 p(x_i|\mathcal{R}, \psi, \omega) dx_i + \gamma_{\mathcal{R}} n \log \sqrt{2\pi} \sigma_{\mathcal{R}} \\
&= \frac{\gamma_{\mathcal{R}} n}{2\sigma_{\mathcal{R}}^2} E\{X^2|\mathcal{R}, \psi_{\mathcal{R}}, \omega\} + \gamma_{\mathcal{R}} n \log \sqrt{2\pi} \sigma_{\mathcal{R}} \\
&= E\{Q^*(\mathcal{R}|\psi_{\mathcal{R}}, \omega)\}.
\end{aligned} \tag{2.19}$$

The claim follows.

Thus far, we have assumed that the posterior probabilities are known but in practice, they must be estimated from the data. This does not, however, affect the consistency of the estimator if the residual distributions can be estimated in a consistent way.

**Theorem 2.3.7** *Consider the estimation of the parameters  $\omega$  for the model  $g$ , as in Theorem 2.3.2 with the only difference being that the posterior probabilities are not known but estimated from the data by fitting the continuous residual model (2.4) using ML estimation. Let the ML estimate for the posterior probability be consistent so that the posterior probability function estimate converges almost surely to the true posterior probability function. Then, the convergence and consistence conditions*

$$(\mathcal{D}|\omega') \xrightarrow[n \rightarrow \infty]{a.s.} E_0\{\tilde{Q}(\mathcal{R}|\psi_{\mathcal{R}}, \omega)\}, \text{ uniform on } \Omega' \tag{2.20}$$

and

$$\hat{\omega}'_n \xrightarrow[n \rightarrow \infty]{P} \omega'_0 \tag{2.21}$$

still hold.

*Proof.* Replace the posterior probabilities in the proof of Theorem 2.3.2 by the estimated counterparts. They converge to the true probabilities uniformly and almost surely since the residual density model is continuous for almost all  $x$  and the parameter space  $\Omega'$  closed and bounded by assumption. The statement follows.

A common result is for the maximum likelihood estimator given Gaussian noise model to be equivalent to the least squares solution. For the maximum likelihood

robust estimator, we obtain a similar result by dropping the positive constants in (2.5).

**Corollary 2.3.8** *The maximum likelihood robust estimator  $\omega'^*$  is equivalent with the weighted least squares solution to*

$$\min_{\omega} \sum_i P(\mathcal{R}|X_i, \psi, \omega) X_i^2, \quad (2.22)$$

where the weights are the a posterior probabilities for relevance given by the Bayes rule.

The result is, moreover, intuitively pleasing. In the outlier free regression assuming Gaussian noise model, the cost of each residual is the square of its value. Here the cost remains the square, but multiplied by the *probability of the cost*! If the noise model for the relevant component were a Laplace distribution, a similar derivation would yield the following result, since 1-norm is the inherent cost function for the exponential distribution.

**Corollary 2.3.9** *Let the relevant residual obey Laplace or double exponential distribution, whereby the maximum likelihood robust estimator is equivalent with solution of the weighted minimization*

$$\min_{\omega} \sum_i P(\mathcal{R}|X_i, \psi, \omega) |X_i|. \quad (2.23)$$

## 2.4 Necessary Conditions

In this section, we will derive the necessary conditions the maximum likelihood robust estimator. Henceforth, we will assume that  $\Psi \subset \mathbb{R}^k$  and  $\Omega \subset \mathbb{R}^l$ . We will also assume, without stating explicitly, that the conditions in Theorems 2.3.2, 2.3.4, and 2.3.7 hold. In practice, these conditions ensure that the solution is unique, and both the outlier-free MLE for the model  $g$  and the MLE for the residual density model are consistent. Now, the objective is to minimize the cost function (2.5) when the posterior probabilities are unknown but the residual models  $p(x|\mathcal{R}, \psi, \omega)$  and  $p(x|\mathcal{F}, \psi, \omega)$  are given.

**Definition 2.4.1 (Problem)** *Compute the maximum likelihood robust estimate  $\hat{\omega}$  for the model  $g$  given data  $\mathcal{D}$  and residual models  $p(x|\mathcal{R}, \psi, \omega)$  and  $p(x|\mathcal{F}, \psi, \omega)$  by minimizing*

$$(\mathcal{D}|\psi, \omega) = \frac{1}{2\sigma_{\mathcal{R}}^2 \gamma_{\mathcal{R}} n} \sum_{i=1}^n P(\mathcal{R}|x_i, \psi, \omega) x_i^2 + \log \sqrt{2\pi} \sigma_{\mathcal{R}}, \quad (2.24)$$

where  $x_i$  is the model residual for the  $i^{\text{th}}$  measurement with the parameters  $\omega \in \Omega$ .

Here, the residual parameter vector  $\psi = (\gamma_{\mathcal{R}}, \psi_{\mathcal{R}}, \psi_{\mathcal{F}}) \in \Psi$  must be estimated from the observed residual to express the posterior probabilities. This can be achieved using maximum likelihood estimation, minimizing the log-likelihood with respect to the residual parameters

$$\psi|_{\omega}^* = \arg \min_{\psi} (\mathcal{D}|\psi, \omega). \quad (2.25)$$

A necessary condition for the solution is that the gradient of the log-likelihood with respect to  $\psi$  vanishes. The entire problem is therefore a constrained optimization problem

$$\min_{\psi, \omega} (\mathcal{D}|\psi, \omega) \quad (2.26)$$

with subject to

$$\frac{\partial (\mathcal{D}|\psi, \omega)}{\partial \psi} = \mathbf{0}. \quad (2.27)$$

By using Lagrange multipliers for equality constraints, we obtain the following set of nonlinear equations

$$\begin{aligned} \frac{\partial}{\partial \omega} + \sum_j \lambda_j \frac{\partial^2}{\partial \omega \partial \psi_j} &= \mathbf{0} \\ \frac{\partial}{\partial \psi} + \sum_j \lambda_j \frac{\partial^2}{\partial \psi \partial \psi_j} &= \mathbf{0} \\ \frac{\partial}{\partial \psi} &= \mathbf{0}, \end{aligned} \quad (2.28)$$

where  $\lambda_j$  is the  $j^{\text{th}}$  Lagrange multiplier. We thus have in total  $2k + l$  equations and as many unknowns. However, rather than attempting to solve the above set of equations directly, we will propose a more practical EM algorithm.

## 2.5 EM Solution

Since the solving of the maximum likelihood robust estimator directly from the non-linear system of equations is difficult, we propose an alternative method based on the expectation maximization (EM) algorithm (Dempster, Laird and Rubin 1977). In general, the EM algorithm involves the complete data  $\mathbf{y} = \{y_i\}$ , which are not available, and incomplete data  $\mathbf{x} = \{x_i\}$ , which have been observed. By computing the expected complete data likelihood, it is, however, possible to find successive estimates such that at each iteration the observed data likelihood  $L(\mathbf{x}|\omega')$  has a larger value (see e.g. Srinath, Rajasekaran and Viswanathan 1996).

The EM algorithm generally has two steps that are iterated until convergence. The *expectation step* involves computing the expected complete data log-likelihood given the observed incomplete data and an estimate for the parameters, forming

$$\chi(\omega', \hat{\omega}'_l) = E_{\mathbf{Y}|\mathbf{x}, \hat{\omega}'_l} \{ \log L(\mathbf{y}|\omega') \mid \mathbf{x}, \hat{\omega}'_l \}. \quad (2.29)$$

In the *maximization step*, new parameter estimates are obtained by

$$\max_{\omega'} \chi(\omega', \hat{\omega}'_l). \quad (2.30)$$

When the likelihood function is bounded and  $\chi(\omega'_1, \omega'_2)$  is continuous, convergence to a local maximum of the likelihood  $L(\mathbf{x}|\omega')$  is guaranteed (Wu 1983).

In our problem, the incomplete data involves all the residual observations  $\mathbf{x} = \{x_i\}$ . Let  $u_i$  be a variable that indicates whether the residual  $x_i$  is a realization from the correct or false residuals. Thus, the corresponding random variable  $U_i$  is either 1 or 0 with probabilities  $\gamma_{\mathcal{R}}$  or  $1 - \gamma_{\mathcal{R}}$ , respectively. We consider the complete data to be  $\mathbf{y} = \{x_i, u_i\}$  with

$$L(\mathbf{y}|\omega') = \prod_{i|x_i \in \mathcal{R}} p(x_i|\mathcal{R}, \omega') = \prod_{i=1}^n (u_i \quad 1 - u_i) \begin{pmatrix} p(x_i|\mathcal{R}, \omega') \\ 1 \end{pmatrix}. \quad (2.31)$$

Hence, the E-step is

$$\begin{aligned} \chi(\omega', \hat{\omega}'_l) &= E \{ \log L(\mathbf{Y}|\omega') \mid \mathbf{x}, \hat{\omega}'_l \} = E \left\{ \sum_i U_i \log p(x_i|\mathcal{R}, \omega') \mid \mathbf{x}, \hat{\omega}'_l \right\} \\ &= \sum_i E \{ U_i \mid \mathbf{x}, \hat{\omega}'_l \} \log p(x_i|\mathcal{R}, \omega') \\ &= - \sum_i P(\mathcal{R}|x_i, \hat{\omega}'_l) \left( \frac{x_i^2}{2\sigma_{\mathcal{R}}^2} + \log \sqrt{2\pi}\sigma_{\mathcal{R}} \right). \end{aligned} \quad (2.32)$$

In the M-step, we thus obtain new estimates for the model parameters by maximizing  $\chi(\omega', \hat{\omega}'_l)$  with respect to  $\omega$ , or equivalently

$$\min_{\omega} \sum_i P(\mathcal{R}|x_i, \hat{\omega}'_l) x_i^2, \quad (2.33)$$

where the positive constants have been dropped (compare to (2.22)). Normally,  $\chi(\omega', \hat{\omega}'_l)$  should be maximized with respect to all parameters  $\omega'$  but here the complete data likelihood does not restrict the residual parameters of the outlying observations. Therefore, to update posterior probability estimates, we must fit the residual model to the data using the ML estimation<sup>4</sup>. In total, we get the algorithm summarized in Algorithm 1. Its implementation is discussed in Section 2.7.

<sup>4</sup>Since this step is added, the algorithm does not represent the EM algorithm in the standard form. We, however, regard it as an EM algorithm because of the way it has been derived.



---

**Algorithm 1** EM Solution for the Maximum Likelihood Robust Estimator

---

1. Compute an initial estimate for the parameters  $\omega$  of the model.
  2. Use the ML estimation to compute estimates for the residual density parameters  $\psi$ , given  $\hat{\omega}_l$ . Update the a posteriori probability estimates  $P(\mathcal{R}|x_i, \hat{\psi}_l, \hat{\omega}_l)$  by using the Bayes rule.
  3. Compute a new estimate for the parameters  $\omega$  by minimizing the squared residuals weighted by the new a posteriori probabilities. Iterate Steps 2 and 3 until the model parameter estimates converge.
- 

## 2.6 Unknown Residual Models

In applications, the actual form of the outlier distribution is rarely known. In robust regression, this fact is usually ignored by *ad hoc* propositions such as M-estimators where it is simply assumed that the total residual distribution is symmetric and has long tails, and the desired density model is *fixed in advance*. A propos, certain adaptive, efficient procedures were developed in the 70's (Hogg 1974, Beran 1974, Stone 1975, Sacks 1975), but their applicability is limited because of the assumption that the total residual is symmetric. As their performance under asymmetry has been unclear and unexplored, they have been excluded from the definition of robust procedures (Huber 1981). Recall that, from our point of view, symmetry is a senseless assumption in general.

Since we assume that good residuals are normally distributed, we model an unknown outlier residual using a mixture of Gaussians. The total residual distribution is

$$p(x|\omega') = \sum_{i=0}^{m-1} \gamma_i p(x|\mathcal{D}_i, \mu_i, \sigma_i, \omega), \quad (2.34)$$

where  $\mathcal{R} = \mathcal{D}_0$ ,  $\mathcal{F} = \mathcal{D}_1 \cup \dots \cup \mathcal{D}_{m-1}$ ,  $\gamma_{\mathcal{R}} = \gamma_0 = 1 - \sum_{i=1}^{m-1} \gamma_i$ . The mixture of Gaussians is a natural choice because, on this basis, we may obtain a clear interpretation for the kernels. For instance, if the regression problem is about motion estimation from several scenes that actually contain several independently moving objects, the other kernels may describe the residuals of the *other* motion clusters. It is therefore more a matter of problem dependent interpretation of which kernel is the “right one”.

A crucial point in modeling the residual distribution using the mixture of Gaussians is how to determine the number of kernels. Since the ML method favors models of ever increasing complexity, we need a means of penalizing models that are overly complex. The complexity control should therefore identify when there

are no outliers in the data, i.e., when only a single Gaussian suffices. The complexity control can be performed in several ways.

The classic Akaike’s Information Criterion (AIC) (Akaike 1977) is derived from Taylor’s polynomials of the mean expected log-likelihood and log-likelihood (see e.g. Sakamoto, Ishiguro and Kitagawa 1986). It is equivalent to minimizing

$$\text{AIC}(\mathcal{D}|\hat{\omega}') = -(\mathcal{D}|\hat{\omega}') + k, \quad (2.35)$$

where  $\hat{\omega}'$  is the maximum likelihood estimate for the parameters and  $k = \dim \Psi$ . The residual parameters in  $\omega'$  are assumed independent.

Another criterion is known as the ‘Bayesian information criterion’ (BIC) (Schwarz 1978), which was derived using Bayesian formalism. The penalty term was found to be  $\frac{k}{2} \log n$  instead of  $k$ , where  $n$  is the number of observations, but the criterion was derived only for the Koopman–Darmois family of distributions. However, Rissanen arrived at the same asymptotic solution using the minimum description length (MDL) principle (Rissanen 1978, Rissanen 1983), but the result can be considered an extension to AIC and BIC for *any sufficiently smooth family of distributions*, assuming that the MLE is consistent in the family. Here, we thus denote

$$\text{MDL}(\mathcal{D}|\hat{\omega}') = -(\mathcal{D}|\hat{\omega}') + \frac{k}{2} \log n. \quad (2.36)$$

Vapnik’s theory of structural risk minimization (SRM) provides another perspective on complexity control (Vapnik 2000). It is based on the statistical analysis of the rate of convergence of an empirical process where the complexity of a model is determined by selecting the model that minimizes the bound on the risk functional. The MDL principle is somewhat similar to the SRM principle but it is philosophically very different. Rissanen himself argues (private communication) that, in order to justify the risk minimization principle, one should show the convergence of the risk functional. In fact, the MDL philosophy goes one step further by stating that there is no such thing as a “true model” (Rissanen 1989).

As the residual form is unknown, we apply the MDL principle to complexity control or to find the number of kernels  $m$  in the Gaussian mixture model according to (2.36), when  $k = 3m - 1$ . The final algorithm is summarized in Algorithm 2. However, because ML is inconsistent for general Gaussian mixtures, we must make it consistent by restricting the parameter space correctly. This and other implementation issues are discussed in the following section.

## 2.7 Implementation Issues

Fitting a Gaussian mixture model to data is anything but a trivial problem, although we here need to consider the 1D residual only. First, because of the

---

**Algorithm 2** Computation of the MLRE with Unknown Residual Models

---

1. Compute an initial estimate for the parameters  $\omega$  of the model.
  2. Determine the ML estimates for the residual likelihood given  $\omega$  for every  $m = 1, \dots, M$  using the stochastic EM algorithm (see Section 2.7). Select the number of kernels  $\hat{m}$  that maximizes (2.36) and store the corresponding residual parameter estimates  $\hat{\psi}$ .
  3. Compute the EM solution for the model parameters  $\omega$  by Algorithm 1 using the  $\hat{m}$  number of kernels and  $\hat{\psi}$  as an initial estimate for the residual parameters.
- 

local minima, the EM solution is found to depend strongly on the initial estimate for the parameters. A mere single run on EM algorithm with a random initial point performs rather poorly (Biernacki, Celeux and Govaert 2000). Second, the mixture model is inconsistent since the likelihood of the mixture tends to infinity as the variance of a kernel centered at a data point approaches zero (see e.g. Vapnik 2000).

In robust regression, we may assume that each residual value has the *true* classification of correct or false even though it is not possible to determine this by examining the residual values. In determining the maximum of the mixture likelihood, we equivalently minimize

$$(\mathcal{D}|\omega') = - \sum_{i=1}^n \log p(x_i|\omega'), \quad (2.37)$$

it would therefore be natural to consider the latent variables

$$z_{ij} = \begin{cases} 1 & \text{if } x_i \text{ belongs to the kernel } j \\ 0 & \text{otherwise} \end{cases} \quad (2.38)$$

rather than trying to solve it directly.

If we knew the values of the latent variables, the solving of the mixture parameters would be simple since we could compute the sample means, variances, and prior probabilities for each group individually, and thereafter evaluate the mixture likelihood. In fact, there are  $m^n$  different classifications in total<sup>5</sup>; thus, at least in principle, we would find the latent variable solution that minimizes (2.37) by simply testing every possible classification. Before the time of quantum computers we must, however, try something else.

---

<sup>5</sup>If different group labelings are excluded, the number of classifications is  $\frac{m^n}{m!}$ .

Previously, Fraley and Raftery (1998) used agglomerative model-based hierarchical clustering to solve the initial values for the latent variables as the final *clustering* was achieved by the EM algorithm. Classification EM (CEM) (Celeux and Govaert 1992), of which the standard k-means is one form, is a variant of the traditional EM algorithm, where the latent variables are determined by a discrete classification in the E-step. In the stochastic EM (SEM) (Celeux and Diebolt 1985), the latent variables are simulated by drawing the  $m$  values using the posterior probabilities of each group in expectation step. Moreover, Biernacki et al. (2000) showed that local minima may be effectively avoided by using CEM or SEM as an appropriate starting point for the EM algorithm.

We use the SEM algorithm here since it should be more exact to our problem than CEM. This is because we must deal with mixtures with overlapping kernels, and it is hence impossible to classify the residuals to good and bad simply by examining the residual values. The implementation of SEM concerns the following details.

First, we need initial estimates for the mixture parameters  $\psi$  given the residual observations  $x_1, \dots, x_n$ . Let  $a = \min x_i$  and  $b = \max x_i$ . We initialize the parameters as

$$\begin{aligned} \gamma_0 &= \frac{1}{2}, \quad \mu_0 = 0, \quad \sigma_0 = \frac{1}{20}(b - a), \\ \gamma_j &= \frac{1}{2(m-1)}, \quad \mu_j = a + \frac{(b-a)(j - \frac{1}{2})}{m-1}, \quad \sigma_j = \frac{b-a}{m-1}, \quad j = 1, \dots, m-1, \end{aligned} \quad (2.39)$$

where  $m$  is the number of kernels in the mixture model. The initial mixture density then resembles a long tailed M-estimator density.

In the second step of SEM, one simulates the class distributions by drawing the class label of each observation randomly from the residual posterior distribution. That is, the class label of the residual  $x_i$  is set to  $\mathcal{D}_j$  with probability

$$P(\mathcal{D}_j | x_i, \omega') = \frac{\gamma_j p(x_i | \mu_j, \sigma_j)}{\sum_{j'=0}^{m-1} \gamma_{j'} p(x_i | \mu_{j'}, \sigma_{j'})}, \quad (2.40)$$

where  $p$  is a Gaussian density function.

After simulating the class labels, we may compute new estimates for the mixture parameters using sample means and variances. For the class  $j$ , we get

$$\gamma_j = \frac{n_j}{n}, \quad \mu_j = \frac{1}{n_j} \sum_{i|x_i \in \mathcal{D}_j} x_i, \quad \sigma_j = \sqrt{\frac{1}{n_j} \sum_{i|x_i \in \mathcal{D}_j} (x_i - \mu_j)^2}. \quad (2.41)$$

At the  $k^{\text{th}}$  iteration, if the class  $j$  does not yield any observations we, however, set

$$\gamma_j^k = \frac{1}{n}, \quad \mu_j^k = \mu_j^{k-1}, \quad \sigma_j^k = \sigma_j^{k-1}, \quad (2.42)$$

---

**Algorithm 3** SEM Algorithm for Fitting a 1D Gaussian Mixture

---

1. Set the initial estimates for the mixture parameters using (2.39).
  2. Draw class labels for each observation from the posterior distribution (2.40).
  3. Compute the sample estimates for the priors, means, and deviations by (2.41), but require also that  $\sigma_j \geq \delta$ ,  $\forall j$ .
  4. Consider the case of a class obtaining only zero or one observations.
  5. Repeat from Step 2 one hundred times. Store the parameters that give the greatest mixture likelihood for the residual observations.
  6. Repeat from Step 1 ten times. The best estimate is the one that gives the greatest likelihood of all 10 estimates.
  7. Refine the best estimate using the classic EM algorithm with the additional constraint  $\sigma_j \geq \delta$ ,  $\forall j$ .
- 

and finally normalize each  $\gamma_j$  by  $\sum_{j'} \gamma_{j'}$ . Similarly, if there is only one observation in the class  $j$ , we set  $\sigma_j^k = \sigma_j^{k-1}$ . To avoid the inconsistency of the estimator, we yet require that  $\sigma_j \geq \delta$ ,  $\forall j$  by setting the deviation always at least to  $\delta$ . The constant  $\delta$  can also be seen as the “precision” parameter, introduced in Rissanen (1989), which makes the mixture distribution bounded.

As the above procedure is repeated iteratively, it forms a Markov Chain, where the state variables are the mixture parameters. After the “burn-in” period, the parameters should concentrate around a stationary point of the likelihood (Birnacki et al. 2000). To obtain reliable estimates, we take the estimate that maximizes the likelihood of the mixture model after iterating 100 times. To make the procedure yet more reliable, we repeat the estimation 10 times by starting the estimation from the initial situation and taking the estimate that gives the greatest likelihood of all the ten estimates. The entire SEM algorithm is summarized in Algorithm 3.

## 2.8 Summary and Discussion

In this chapter, we have proven that, given a Gaussian noise model for “good” residuals and an independent residual model for outliers, the estimator that minimizes the sum of all squared residuals weighted by the consistent ML estimates

of the posterior probabilities to be correct is a *consistent* estimator for a model's *true* parameter values. This means optimality since the ML estimate of the model is asymptotically achieved regardless of the outliers in the data. The proposed estimator was therefore named as the maximum likelihood robust estimator. The assumptions made are similar to those needed for the consistency of the ordinary ML estimator. The result cannot be generally proven to other robust estimators such as the least-median-of-squares or M-estimators because of, among other things, their assumption of a priori fixed residual distributions.

Where the outlier residual model is known, the maximum likelihood robust estimator can be solved using a simple EM-algorithm; however, the general case of unknown outlier residual models was additionally considered in the chapter. In the latter case, the outlier distribution was proposed to be modeled by a Gaussian mixture model, where the number of kernels is determined using the MDL principle. The function set of Gaussian mixtures is a natural choice since it is a universal approximator. In motion estimation, for example, the outliers may also represent structured noise as in the presence of many motion clusters. Individual kernels in the residual model may therefore involve a certain problem dependent interpretation.

Finally, we may ask to which category the maximum likelihood robust estimator belongs if we follow Rousseeuw and Leroy's (1987) taxonomy of regression diagnostics and robust statistics (see Section 2.1). The method contains features from both classes, and it falls therefore between these categories. As is typical with robust statistics, our method searches for a robust fit without trying to first identify the outliers. On the other hand, the proposed EM algorithm iteratively downgrades certain residual samples in a similar way to the pinpointing and rejecting process of regression diagnostics. In contrast to approaches that attempt to identify outliers, irrespective of into which category they fall, our method gives only a probability for each observation of being correct instead; it is, in fact, the most that one can obtain by simply examining the residual values.

## Chapter 3

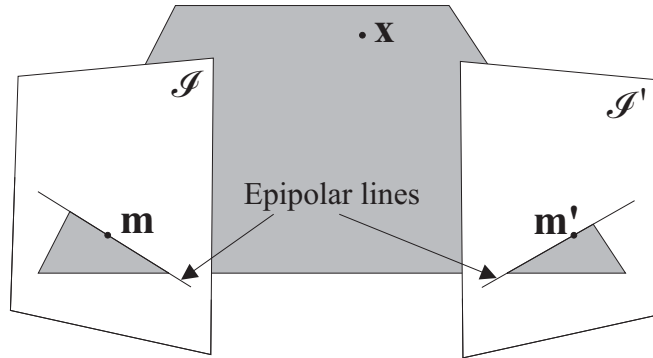
# Two View Geometry Estimation<sup>1</sup>

Matching two images of a single scene is one of the fundamental problems in computer vision. Stereo, motion, and 3D object recognition are all multiple view problems sharing the same geometrical constraint, the epipolar constraint, and numerous techniques have been proposed for its estimation. The methods can be divided into ordinary and outlier-robust methods where some of the ordinary methods are optimal in the sense that they minimize the squared residuals corresponding to a random variable which may be justified to be normally distributed. They fail, however, if there are false point matches in the data. The robust methods attempt to take the false matches into consideration but they violate the Gaussian distribution assumption of the good point residuals, and may even show poor performance under Gaussian noise on the points.

In this chapter, we evaluate the maximum likelihood robust estimator, proposed in the previous chapter, in epipolar geometry estimation. It turns out that, when the affine camera model can be assumed, we are able to build a reasonable form for the outlier distribution. The first part of the chapter thus focuses on the affine epipolar geometry estimation where no automatic model selection for the outlier residuals is needed. The affine camera model is also ideal for an electron microscope (see Chapters 7 and 8). We experiment the general projective case with the assumption of an unknown outlier residual model. In addition, we show how to determine the uncertainty of the epipolar geometry in both affine and projective cases by using all the observed data rather than the (erroneously) relevant classified matching points.

---

<sup>1</sup>This chapter relies on the most recent results of affine (Brandt and Heikkonen 2001*b*) and projective (Brandt 2002*b*) epipolar geometry and its uncertainty estimation, but the first considerations were published in Brandt and Heikkonen (2000*b*) and Brandt and Heikkonen (2000*d*).



**Figure 3.1:** Epipolar geometry involving two images. A three dimensional point  $\mathbf{x}$  and its projections on two image planes form an epipolar plane. The plane intersects the image planes on lines called epipolar lines. Equivalently, when a point in one image is fixed, the real 3D point lies on the back projection ray that lies on the epipolar plane; therefore, the corresponding point in the other image must lie on the corresponding epipolar line.

### 3.1 Introduction

Epipolar geometry refers to the intrinsic projective geometry of two views, and it depends only on the relative pose of the two image planes and internal parameters of the cameras (see Fig. 3.1). Epipolar geometry can be represented by a  $3 \times 3$  matrix known as the fundamental matrix, which is a generalization of the essential matrix (Longuet-Higgins 1981) for uncalibrated cameras. The fundamental matrix  $\mathbf{F}$  is defined by the relation

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0 \quad (3.1)$$

for any matching points pairs  $\mathbf{m} \leftrightarrow \mathbf{m}'$  in two images, represented in homogeneous coordinates. The importance of the fundamental matrix has been stressed, and its role in future applications has been predicted to be crucial. This is supported by its applicability to several difficult problems in computer vision (Luong and Faugeras 1996).

With the full perspective projection camera model, an analytical solution for the fundamental matrix can be achieved with seven point matches because it has only seven degrees of freedom (see e.g. Xu and Zhang 1996). A unique solution is obtained if eight or more point matches in a general position are used; yet the problem becomes over determined when the least squares solution is usually sought. In order to make the eight-point algorithm more accurate and stable, Hartley (1997) has proposed a better way to normalize the matching point coordinates. Other kinds of linear approaches to be mentioned include the Boufama and Mohr's (1995) method, which is based on the concept of virtual parallax, and Ponce and Genc's (1996) linear subspace method.



The linear criterion does not have a clear geometric interpretation when it would be reasonable to assume a normally distributed residual. An optimal method, assuming Gaussian noise in point coordinates, is to minimize the geometrical or reprojection error (Hartley 1993, Hartley 1994, Kanatani 1996). Computationally lighter approximations of this error include the gradient-weighted epipolar errors (Weng, Huang and Ahuja 1989) (for an efficient algorithm, see Chojnacki, Brooks, van den Hengel and Gawley 2000, Chojnacki, Brooks, van den Hengel and Gawley 2002) and the sum of square distance of points and their corresponding epipolar lines (Weng et al. 1989, Luong, Deriche, Faugeras and Papadopoulos 1993). Interpretation and evaluation of the reprojection error vs. its approximations are presented by Zhang (1998b).

The methods presented so far are sensitive to outliers, because, on balance, the square sum over the residuals is minimized. In other words, these approaches correspond to the assumption that the residual is normally distributed, which is clearly not the case if there are outliers in the data. Therefore, certain robust estimators have been used, of which the best known are perhaps the M-estimators and the Least Median of Squares (LMedS) method (see Xu and Zhang 1996). In addition, F-matrix is frequently estimated robustly by the RANSAC (Random Sample Consensus) (Fischler and Bolles 1981) principle (see e.g. Torr and Murray 1993, Torr and Murray 1997).

As reviewed in the previous chapter, M-estimators are based on replacing the squared residual by another function of residuals to make the estimation less sensitive to outliers. Xu and Zhang (1996) reported that, in F-matrix estimation, M-estimators are robust to bad match localizations but not to false matches. Nevertheless, Torr and Murray (1997) obtained good results with a hybrid method that is based on refining the initial estimate at several stages.

The LMedS method is based on nonlinear minimization, where the median of squared residuals is minimized. Accordingly,  $7 \times m$  point correspondences are sampled randomly to achieve  $m$  sets of seven point pairs for which the  $m$  fundamental matrices are estimated separately. The quantity to be minimized is the median residual<sup>2</sup> over all point pairs, i.e.,

$$\min_j \operatorname{median}_i \epsilon_i(\mathbf{F}_j)^2. \quad (3.2)$$

Although good results have been obtained with this method, the efficiency of LMedS has been characterized as poor in the presence of Gaussian noise (Xu and Zhang 1996). To reduce this deficiency, minimization is followed by an M-estimator.

In the RANSAC method, an F-matrix is estimated from several random minimal matching point sets as well. Such an F-matrix is preferred which maximizes the

---

<sup>2</sup>Even though we used  $x$  as the most convenient symbol for residuals in the previous chapter, henceforth we use  $\epsilon$  because we want to reserve the symbol  $x$  for image coordinates.

number of inliers, i.e., the number of data points having smaller residual than a heuristic threshold. The final estimate is computed using only the inliers and minimizing the reprojection error or its approximation. It therefore assumes that the inliers are normally distributed, which is not true yet.

Previously, Torr and Zisserman (2000) proposed a RANSAC like estimator known as MLESAC (Maximum Likelihood Sample Consensus). The objective was to maximize, using random sampling, the likelihood of the mixture of two densities corresponding to relevant and false matches. One problem with this method, as well as with MSAC (M-Estimator Sample Consensus) in the same paper, is its assumption of known standard deviation of the noise; we should search for the minimum variance estimate that becomes meaningless if the noise variance is fixed.

### 3.2 Affine F-Matrix Estimation

With an affine camera model (orthographic, weak perspective, or paraperspective, for details, see Xu and Zhang 1996), the problem of fundamental matrix estimation is slightly simpler because the epipolar equation  $\mathbf{m}^T \mathbf{F} \mathbf{m} = 0$  is linear in image coordinates. It may thus be expanded as

$$\mathbf{u}_i^T \mathbf{f} + f_{33} = 0, \quad (3.3)$$

where  $\mathbf{f} = (f_{13} \ f_{23} \ f_{31} \ f_{32})^T$  and  $\mathbf{u}_i = (x'_i \ y'_i \ x_i \ y_i)^T$ . A minimum of four matched point pairs are sufficient to determine the affine fundamental matrix uniquely. If more matches are used and the reprojection error is minimized, it can be shown (Xu and Zhang 1996) that the solution for  $\mathbf{f}$  is obtained from the following eigenequation as the eigenvector corresponding to the smallest eigenvalue:

$$\mathbf{W} \mathbf{f} - \lambda \mathbf{f} = \mathbf{0}, \quad (3.4)$$

where  $\mathbf{W} = \sum_{i=1}^N (\mathbf{u}_i - \mathbf{u}_0)(\mathbf{u}_i - \mathbf{u}_0)^T$ ,  $\mathbf{u}_0 = \frac{1}{N} \sum_{i=1}^N \mathbf{u}_i$ . The parameter  $f_{33}$  is obtained from

$$f_{33} = -\mathbf{u}_0^T \mathbf{f}. \quad (3.5)$$

Because it is assumed here that the error is normally distributed, the result is similarly sensitive to outliers. To accommodate this, Zhang proposes an affine version<sup>3</sup> of the LMedS-method. In addition, other robust estimators, including RANSAC, are straightforward to implement in the affine case.

It was mentioned earlier that the proposed maximum likelihood robust estimator is based on the fact that the residuals corresponding to the outliers, which are

---

<sup>3</sup>Available from <http://www-sop.inria.fr/robotvis/personnel/zzhang/zzhang-eng.html>.

here regarded as false matches, come from a different distribution than the relevant point matches. For the distribution of the relevant residuals, it is reasonable to assume Gaussian distribution if the geometric error is minimized. The distribution of false matches is generally unknown but, with the affine camera model, it is reasonable to assume it is similarly normal, but with *different* parameters. The rationale lies in the central limit theorem, which states that the distribution of the sum of i.i.d. random variables approaches a normal distribution as the number of variables grows, as explained in the following.

Assuming an affine camera model, the residual of a false match is defined as

$$\epsilon_f = \frac{\mathbf{u}_i^T \mathbf{f} + f_{33}}{\|\mathbf{f}\|_2}, \quad \mathbf{u}_i \in \mathcal{F} \quad (3.6)$$

where  $\mathcal{F}$  is the set of false matches. Because  $\epsilon_f$  can be considered a sum of *four* random variables ( $\mathbf{u}_i \in \mathbb{R}^4$ ), the normal distribution assumption is reasonable (central limit theorem). One could still claim that there are far too few variables to make such a distribution assumption. However, the density function of the sum of random variables is obtained by convolving the densities of the individual random variables. If we assume that the false matches are uniformly distributed to the image planes, which is perhaps the best a priori assumption we can make<sup>4</sup>, the density of the sum would be the B-spline of the degree of four which is, indeed, a very Gaussian like bell function.

We therefore assume that the residuals of mismatches are normally distributed, and consider Algorithm 1 to solve MLRE. According to the following proposition, the minimization of the weighted reprojection error can be computed in closed-form.

**Proposition 3.2.1** *Let the posterior probabilities  $P(\mathcal{R}|\epsilon_i, \omega')$  be fixed. Then, with the affine camera model, the posterior probability weighted minimization of the squared reprojection error, yields the solution of the form (3.4) and (3.5) where*

$$\mathbf{W} = \sum_i P(\mathcal{R}|\epsilon_i, \omega') (\mathbf{u}_i - \mathbf{u}_0)(\mathbf{u}_i - \mathbf{u}_0)^T \quad (3.7)$$

and

$$\mathbf{u}_0 = \frac{\sum_i P(\mathcal{R}|\epsilon_i, \omega') \mathbf{u}_i}{\sum_i P(\mathcal{R}|\epsilon_i, \omega')}. \quad (3.8)$$

*Proof.* This analytical derivation is a generalized version of the derivation presented in Xu and Zhang (1996). The objective here is to minimize the weighted

---

<sup>4</sup>This assumption is slightly pessimistic because, in practice, the false matches seem to distribute bell-like on the image planes, i.e., the Gaussian approximation should work even better.

sum of squared residuals with respect to  $\mathbf{F}$  (whose parameterization is represented by  $\omega$ ), i.e.,

$$C = \sum_{i=1}^N P(\mathcal{R}|\epsilon_i, \omega') \epsilon_i^2 = \sum_{i=1}^N P_i \frac{(\mathbf{u}_i^T \mathbf{f} + f_{33})^2}{\mathbf{f}^T \mathbf{f}}, \quad (3.9)$$

where the short hand notation  $P_i = P(\mathcal{R}|\epsilon_i, \omega')$  is used. Differentiating  $C$  with respect to  $f_{33}$  and setting the result to zero, we have

$$\frac{\partial C}{\partial f_{33}} = 2 \sum_{i=1}^N P_i \frac{\mathbf{u}_i^T \mathbf{f} + f_{33}}{\mathbf{f}^T \mathbf{f}} = 0. \quad (3.10)$$

Therefore, the solution for  $f_{33}$  is

$$f_{33} = -\frac{\sum_{i=1}^N P_i \mathbf{u}_i^T \mathbf{f}}{\sum_{i=1}^N P_i} = -\mathbf{u}_0^T \mathbf{f}, \quad (3.11)$$

where  $\mathbf{u}_0 = \sum_i P_i \mathbf{u}_i / \sum_i P_i$ . Substituting now (3.11) into (3.9), it follows

$$C = \sum_{i=1}^N P_i \frac{(\mathbf{u}_i^T \mathbf{f} - \mathbf{u}_0^T \mathbf{f})^2}{\mathbf{f}^T \mathbf{f}}. \quad (3.12)$$

Let  $\mathbf{W} = \sum_{i=1}^N P_i (\mathbf{u}_i - \mathbf{u}_0)(\mathbf{u}_i - \mathbf{u}_0)^T$ , then it follows

$$C = \sum_{i=1}^N P_i \frac{\mathbf{f}^T (\mathbf{u}_i - \mathbf{u}_0)(\mathbf{u}_i - \mathbf{u}_0)^T \mathbf{f}}{\mathbf{f}^T \mathbf{f}} = \frac{\mathbf{f}^T \mathbf{W} \mathbf{f}}{\mathbf{f}^T \mathbf{f}}. \quad (3.13)$$

Differentiating with respect to  $\mathbf{f}$  we get

$$\frac{\partial C}{\partial \mathbf{f}} = \frac{2\mathbf{W}\mathbf{f}}{\mathbf{f}^T \mathbf{f}} - \frac{2\mathbf{f}^T \mathbf{W} \mathbf{f} \mathbf{f}}{(\mathbf{f}^T \mathbf{f})^2} = 2 \frac{\mathbf{W}\mathbf{f} - C\mathbf{f}}{\mathbf{f}^T \mathbf{f}}. \quad (3.14)$$

Setting the equation to zero, we obtain the eigenequation

$$\mathbf{W}\mathbf{f} - C\mathbf{f} = \mathbf{0}. \quad (3.15)$$

Now, because  $\mathbf{W}$  is symmetric and positive semi-definite, it has 4 real non-negative eigenvalues and associated eigenvectors. Since we seek to minimize  $C$ , we choose the minimal eigenvalue and the associated eigenvector.

The EM implementation (see Section 2.5) of the maximum likelihood robust estimator for affine fundamental matrix is now summarized in Algorithm 4. Independent of our work, a somewhat similar EM algorithm has been used in the projective F-matrix estimation by Torr and Murray (1997) (see also Zhuang, Wang and Zhang 1992), and good results were reported. However, in their work the matches are *strictly classified* to inliers and outliers, after which iterative M-estimators are used for the inliers. Because of the classification, their solution does not converge to the ML estimate or, to be precise, is not consistent although the distribution assumptions were correct.

---

**Algorithm 4** Implementation of the MLRE for the Affine F-Matrix Estimation

---

1. Compute an initial estimate for the affine fundamental matrix e.g. with the affine LMedS.
  2. Use the ML estimation to compute the parameters for the density functions and a priori probabilities of the reprojection error of the relevant and false matches. Update the a posteriori probability estimates  $P(\mathcal{R}|\epsilon_i, \omega')$  by using the Bayes rule.
  3. Compute a new estimate for  $\mathbf{f}$  by solving the eigenvector corresponding to the smallest eigenvalue of  $\mathbf{W}$  (3.7). Compute the element  $f_{33}$  from (3.5) and (3.8). Iterate Steps 2 and 3 until the fundamental matrix estimate converges.
- 

### 3.3 Uncertainty of the Affine F-matrix

The present approaches for the F-matrix uncertainty estimation rely on the assumption that all the point matches used in computing the covariance matrix are relevant. In order to accomplish this, the matches must first be classified as relevant and false. However, as reported in the previous chapter, classification without error is impossible because the residual densities overlap. Indeed, measuring uncertainty in this way is questionable since the *ad hoc* inlier threshold determines the largest possible error. The classification additionally violates the assumption that the relevant residuals are normally distributed.

In this section, we demonstrate that by following the proposed probabilistic principle, the covariance estimation of the affine F-matrix can be generalized from the original work of Shapiro, Zisserman and Brady (1994) and Shapiro and Brady (1995). In contrast to their work, we allow the presence of outliers. In general, the analytical first-order covariance approximation has been reported to be quite good when the noise level in data points is moderate, or the standard deviation is under two pixels (Xu and Zhang 1996, Csurka, Zeller, Zhang and Faugeras 1997).

**Proposition 3.3.1** *Consider the affine fundamental matrix estimation from outlier contaminated point matches. Let  $P(\mathcal{R}|\epsilon_i, \omega')$  be the posterior probability for the match  $i$  obtained by the maximum likelihood robust estimator. Then, the first-order estimate for the covariance matrix  $\mathbf{C}_{\mathbf{f}}$  for the parameterization  $\mathbf{f} = (f_{13} \ f_{23} \ f_{31} \ f_{32})^T$  is given by*

$$\mathbf{C}_{\mathbf{f}} \simeq \sigma^2 \mathbf{Q} \left( \sum_i P(\mathcal{R}|\epsilon_i, \omega')^2 (\mathbf{u}_i - \mathbf{u}_0)(\mathbf{u}_i - \mathbf{u}_0)^T \right) \mathbf{Q}^T, \quad (3.16)$$

where

$$\mathbf{Q} = - \sum_{k=2}^4 \frac{\mathbf{q}_k \mathbf{q}_k^T}{\lambda_k} \quad (3.17)$$

whereas  $\lambda_k$  and  $\mathbf{q}_k$  are the  $k^{\text{th}}$  largest eigenvalue and the associated eigenvector of  $\mathbf{W}$  from (3.7), respectively.

*Proof.* Without loss of generality, we may assume that both relevant and false matches are perturbed by additive Gaussian noise. Let  $\bar{\mathbf{u}}_i$  be the true, noise free 4D-coordinates of the match  $i$  and the corresponding noise vector  $\delta \mathbf{u}_i$ . Now,

$$\mathbf{u}_i = \bar{\mathbf{u}}_i + \delta \mathbf{u}_i. \quad (3.18)$$

It is further assumed that the noise is uncorrelated and has zero mean and equal variance in all the four components. Briefly, we require  $E\{\delta \mathbf{u}_i\} = 0$  and

$$\mathbf{C}_{\mathbf{u}} = E\{\delta \mathbf{u}_i \delta \mathbf{u}_i^T\} = \sigma^2 \mathbf{I}_4. \quad (3.19)$$

From the independence, it follows

$$E\{\delta \mathbf{u}_i \delta \mathbf{u}_j^T\} = \delta_{ij} \mathbf{C}_{\mathbf{u}}, \quad (3.20)$$

where  $\delta_{ij}$  is the Kronecker delta product.

To clarify future calculations, let us define  $\mathbf{v}_i = \sqrt{P_i}(\mathbf{u}_i - \mathbf{u}_0)$ , where  $P_i$  is a shorthand notation for  $P(\mathcal{R}|\epsilon_i, \omega')$  and  $\mathbf{u}_0 = \sum_i P_i \mathbf{u}_i / \sum_i P_i$  as in (3.12). Now,  $\mathbf{v}_i$  is divided into its noise free and noise components. By substituting (3.18) and arranging terms, we get

$$\begin{aligned} \mathbf{v}_i &= \sqrt{P_i} \left( \bar{\mathbf{u}}_i - \underbrace{\frac{\sum_i P_i \bar{\mathbf{u}}_i}{\sum_i P_i}}_{\triangleq \bar{\mathbf{u}}_0} \right) + \sqrt{P_i} \left( \delta \mathbf{u}_i - \underbrace{\frac{\sum_i P_i \delta \mathbf{u}_i}{\sum_i P_i}}_{\triangleq \delta \mathbf{u}_0} \right) \\ &= \sqrt{P_i} (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_0) + \sqrt{P_i} (\delta \mathbf{u}_i - \delta \mathbf{u}_0) \\ &= \bar{\mathbf{v}}_i + \delta \mathbf{v}_i, \end{aligned} \quad (3.21)$$

where  $\bar{\mathbf{v}}_i = \sqrt{P_i}(\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_0)$  and  $\delta \mathbf{v}_i = \sqrt{P_i}(\delta \mathbf{u}_i - \delta \mathbf{u}_0)$ . Now, it follows

$$\begin{aligned} E\{\delta \mathbf{v}_i \delta \mathbf{v}_j^T\} &= E\{\sqrt{P_i P_j} (\delta \mathbf{u}_i - \delta \mathbf{u}_0) (\delta \mathbf{u}_j - \delta \mathbf{u}_0)^T\} \\ &= \sqrt{P_i P_j} (E\{\delta \mathbf{u}_i \delta \mathbf{u}_j^T\} - E\{\delta \mathbf{u}_i \delta \mathbf{u}_0^T\} - E\{\delta \mathbf{u}_0 \delta \mathbf{u}_j^T\} + E\{\delta \mathbf{u}_0 \delta \mathbf{u}_0^T\}) \\ &= \sqrt{P_i P_j} \left( E\{\delta \mathbf{u}_i \delta \mathbf{u}_j^T\} - E\{\delta \mathbf{u}_i \frac{\sum_i P_i \delta \mathbf{u}_i^T}{\sum_i P_i}\} - E\{\frac{\sum_i P_i \delta \mathbf{u}_i}{\sum_i P_i} \delta \mathbf{u}_j^T\} + \right. \\ &\quad \left. + E\{\frac{\sum_i P_i \delta \mathbf{u}_i}{\sum_i P_i} \frac{\sum_i P_i \delta \mathbf{u}_i^T}{\sum_i P_i}\} \right) \\ &= \sqrt{P_i P_j} \left( \delta_{ij} \mathbf{C}_{\mathbf{u}} - \frac{P_i}{\sum_i P_i} \mathbf{C}_{\mathbf{u}} - \frac{P_j}{\sum_j P_j} \mathbf{C}_{\mathbf{u}} + \frac{\sum_i P_i^2}{(\sum_i P_i)^2} \mathbf{C}_{\mathbf{u}} \right) \end{aligned}$$

$$= \sqrt{P_i P_j} \left( \delta_{ij} - \frac{P_i + P_j}{\sum_i P_i} + \frac{\sum_i P_i^2}{(\sum_i P_i)^2} \right) \mathbf{C}_u. \quad (3.22)$$

Let us next define  $\mathbf{V} = (\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_N)$  and  $\delta\mathbf{V} = (\delta\mathbf{v}_1 \ \delta\mathbf{v}_2 \ \cdots \ \delta\mathbf{v}_N)$ . Now, using the definition of  $\mathbf{W}$  from (3.13) it follows that  $\mathbf{W} = \mathbf{V}\mathbf{V}^T$ , thus

$$\mathbf{W} = (\bar{\mathbf{V}} + \delta\mathbf{V}) (\bar{\mathbf{V}} + \delta\mathbf{V})^T = \underbrace{\bar{\mathbf{V}}\bar{\mathbf{V}}^T}_{\bar{\mathbf{W}}} + \underbrace{\bar{\mathbf{V}}\delta\mathbf{V}^T + \delta\mathbf{V}\bar{\mathbf{V}}^T + \delta\mathbf{V}\delta\mathbf{V}^T}_{\delta\mathbf{W}}. \quad (3.23)$$

By dividing  $\mathbf{W}$  to its noise free and noise dependent components  $\bar{\mathbf{W}}$  and  $\delta\mathbf{W}$ , and ignoring the second order term, we may write

$$\delta\mathbf{W} \simeq \bar{\mathbf{V}}\delta\mathbf{V}^T + \delta\mathbf{V}\bar{\mathbf{V}}^T. \quad (3.24)$$

Let the eigenvector of  $\mathbf{W}$  be  $\mathbf{q}_k$  as  $k = 1, \dots, 4$  which are ordered in the ascending order according to the corresponding eigenvalues. Using the fact that  $\delta\mathbf{W}$  is symmetric and in the noise free case the smallest eigenvalue  $\bar{\lambda}_1 = 0$ , the first-order perturbation  $\delta\bar{\mathbf{q}}_1$  of the first eigenvector  $\bar{\mathbf{q}}_1$  may be derived from the Taylor expansion (Golub and Loan 1989), which yields

$$\delta\bar{\mathbf{q}}_1 \simeq - \sum_{k=2}^4 \frac{\bar{\mathbf{q}}_k^T \delta\mathbf{W} \bar{\mathbf{q}}_1 \bar{\mathbf{q}}_k}{\bar{\lambda}_k} = \bar{\mathbf{Q}} \delta\mathbf{W} \bar{\mathbf{q}}_1, \quad (3.25)$$

where

$$\bar{\mathbf{Q}} = - \sum_{k=2}^4 \frac{\bar{\mathbf{q}}_k \bar{\mathbf{q}}_k^T}{\bar{\lambda}_k}. \quad (3.26)$$

Substituting (3.24) to (3.25), we get

$$\delta\bar{\mathbf{q}}_1 \simeq \bar{\mathbf{Q}} (\bar{\mathbf{V}}\delta\mathbf{V}^T + \delta\mathbf{V}\bar{\mathbf{V}}^T) \bar{\mathbf{q}}_1 = \bar{\mathbf{Q}} \sum_{i=1}^N (\bar{\mathbf{v}}_i \delta\mathbf{v}_i^T \bar{\mathbf{q}}_1 + \underbrace{\delta\mathbf{v}_i \bar{\mathbf{v}}_i^T \bar{\mathbf{q}}_1}_{=0}), \quad (3.27)$$

where the second term is zero because  $\bar{\mathbf{v}}_i^T \bar{\mathbf{q}}_1$  approaches zero for all  $i$  when the noise variance tends to zero. This is due to the fact that when noise approaches zero, the estimated a posteriori probabilities approach unity for the relevant matches and zero for false matches. In other words, as noise goes to zero, the relevant distribution becomes the Dirac delta function, and the a posteriori probability approaches the Kronecker delta function, being one for relevant matches. For the relevant matches, the term is also zero because  $\bar{\mathbf{v}}_i^T \bar{\mathbf{q}}_1 = 0$  is equivalent that (3.3) holds, i.e., the residual vanishes.

Thus, we may approximate the covariance of  $\bar{\mathbf{q}}_1$  as follows:

$$\begin{aligned}
\mathbf{C}_{\bar{\mathbf{q}}_1} &= \mathbb{E}\{\delta\bar{\mathbf{q}}_1\delta\bar{\mathbf{q}}_1^T\} \simeq \bar{\mathbf{Q}}\mathbb{E}\left\{\left(\sum_{i=1}^N \bar{\mathbf{v}}_i\delta\mathbf{v}_i^T\bar{\mathbf{q}}_1\right)\left(\sum_{j=1}^N \bar{\mathbf{v}}_j\delta\mathbf{v}_j^T\bar{\mathbf{q}}_1\right)^T\right\}\bar{\mathbf{Q}}^T \\
&= \bar{\mathbf{Q}}\mathbb{E}\left\{\sum_{i=1}^N \sum_{j=1}^N \bar{\mathbf{v}}_i\delta\mathbf{v}_i^T\bar{\mathbf{q}}_1\bar{\mathbf{q}}_1^T\delta\mathbf{v}_j\bar{\mathbf{v}}_j^T\right\}\bar{\mathbf{Q}}^T \\
&= \bar{\mathbf{Q}}\left(\sum_{i=1}^N \sum_{j=1}^N \bar{\mathbf{v}}_i\bar{\mathbf{v}}_j^T\bar{\mathbf{q}}_1^T\mathbb{E}\{\delta\mathbf{v}_i\delta\mathbf{v}_j^T\}\bar{\mathbf{q}}_1\right)\bar{\mathbf{Q}}^T.
\end{aligned} \tag{3.28}$$

Substituting (3.22) and noting that  $\bar{\mathbf{q}}_1^T\mathbb{E}\{\delta\mathbf{v}_i\delta\mathbf{v}_j^T\}\bar{\mathbf{q}}_1 = \mathbb{E}\{\delta\mathbf{v}_i\delta\mathbf{v}_j^T\}$ , which follows from (3.19) and  $\bar{\mathbf{q}}_1^T\bar{\mathbf{q}}_1 = 1$ , gives

$$\begin{aligned}
\mathbf{C}_{\bar{\mathbf{q}}_1} &\simeq \sigma^2\bar{\mathbf{Q}}\left(\sum_{i=1}^N \sum_{j=1}^N \sqrt{P_iP_j}\left(\delta_{ij} - \frac{P_i + P_j}{\sum_i P_i} + \frac{\sum_i P_i^2}{(\sum_i P_i)^2}\right)\bar{\mathbf{v}}_i\bar{\mathbf{v}}_j^T\bar{\mathbf{Q}}^T\right) \\
&= \sigma^2\bar{\mathbf{Q}}\left(\sum_{i=1}^N P_i\bar{\mathbf{v}}_i\bar{\mathbf{v}}_i^T\right)\bar{\mathbf{Q}}^T - \sigma^2\bar{\mathbf{Q}}\left(\sum_{i=1}^N \frac{P_i^{3/2}}{\sum_i P_i}\bar{\mathbf{v}}_i \underbrace{\sum_{j=1}^N \sqrt{P_j}\bar{\mathbf{v}}_j^T}_{=0}\right)\bar{\mathbf{Q}}^T + \\
&\quad - \sigma^2\bar{\mathbf{Q}}\left(\underbrace{\sum_{i=1}^N \sqrt{P_i}\bar{\mathbf{v}}_i}_{=0} \sum_{j=1}^N \frac{P_j^{3/2}}{\sum_i P_i}\bar{\mathbf{v}}_j^T\right)\bar{\mathbf{Q}}^T + \\
&\quad + \sigma^2 \frac{\sum_i P_i^2}{(\sum_i P_i)^2}\bar{\mathbf{Q}}\left(\underbrace{\sum_{i=1}^N \sqrt{P_i}\bar{\mathbf{v}}_i}_{=0} \underbrace{\sum_{j=1}^N \sqrt{P_j}\bar{\mathbf{v}}_j^T}_{=0}\right)\bar{\mathbf{Q}}^T.
\end{aligned} \tag{3.29}$$

The above equation consists of the noise free components, which are not generally known. Correspondingly, as in Shapiro and Brady (1995) and Weng et al. (1989), the noise free components can, however, be approximated by the corresponding noisy values. This is justified by the fact that, if one substitutes  $\bar{\mathbf{V}} = \mathbf{V} - \delta\mathbf{V}$  in the relevant equations, the terms in  $\delta\mathbf{V}$  disappear in the first-order approximations, which allows  $\bar{\mathbf{V}}$  simply to be interchanged with  $\mathbf{V}$ . We hence obtain the covariance matrix approximation for  $\mathbf{f}$  from

$$\mathbf{C}_{\mathbf{f}} \simeq \sigma^2\mathbf{Q}\left(\sum_i P(\mathcal{R}|\epsilon_i, \omega')^2(\mathbf{u}_i - \mathbf{u}_0)(\mathbf{u}_i - \mathbf{u}_0)^T\right)\mathbf{Q}^T, \tag{3.30}$$

where

$$\mathbf{Q} = -\sum_{k=2}^4 \frac{\mathbf{q}_k\mathbf{q}_k^T}{\lambda_k}. \tag{3.31}$$



## 3.4 Affine Experiments

The maximum likelihood robust estimator for the affine fundamental matrix has been experimented with both synthetic and real data. In order to evaluate the F-matrix estimates, they are compared to the true or ground truth F-matrices. Because the elements of the F-matrix do not all have equal variances, the Frobenius norm of the matrix difference would not yield a true picture of the similarity. Therefore, we use Zhang's method<sup>5</sup> (Zhang 1998a) in measuring the difference between the ground truth and estimated F-matrix. The difference is here denoted by  $\Delta\mathbf{F}$ . The comparison is based on sampling points from images and the corresponding epipolar lines randomly, and measuring the distance between points and the epipolar lines computed crosswise between the two F-matrices (see Zhang 1998a). We thus have a measure of *generalization* for the F-matrix estimation techniques.

### 3.4.1 Synthetic Data

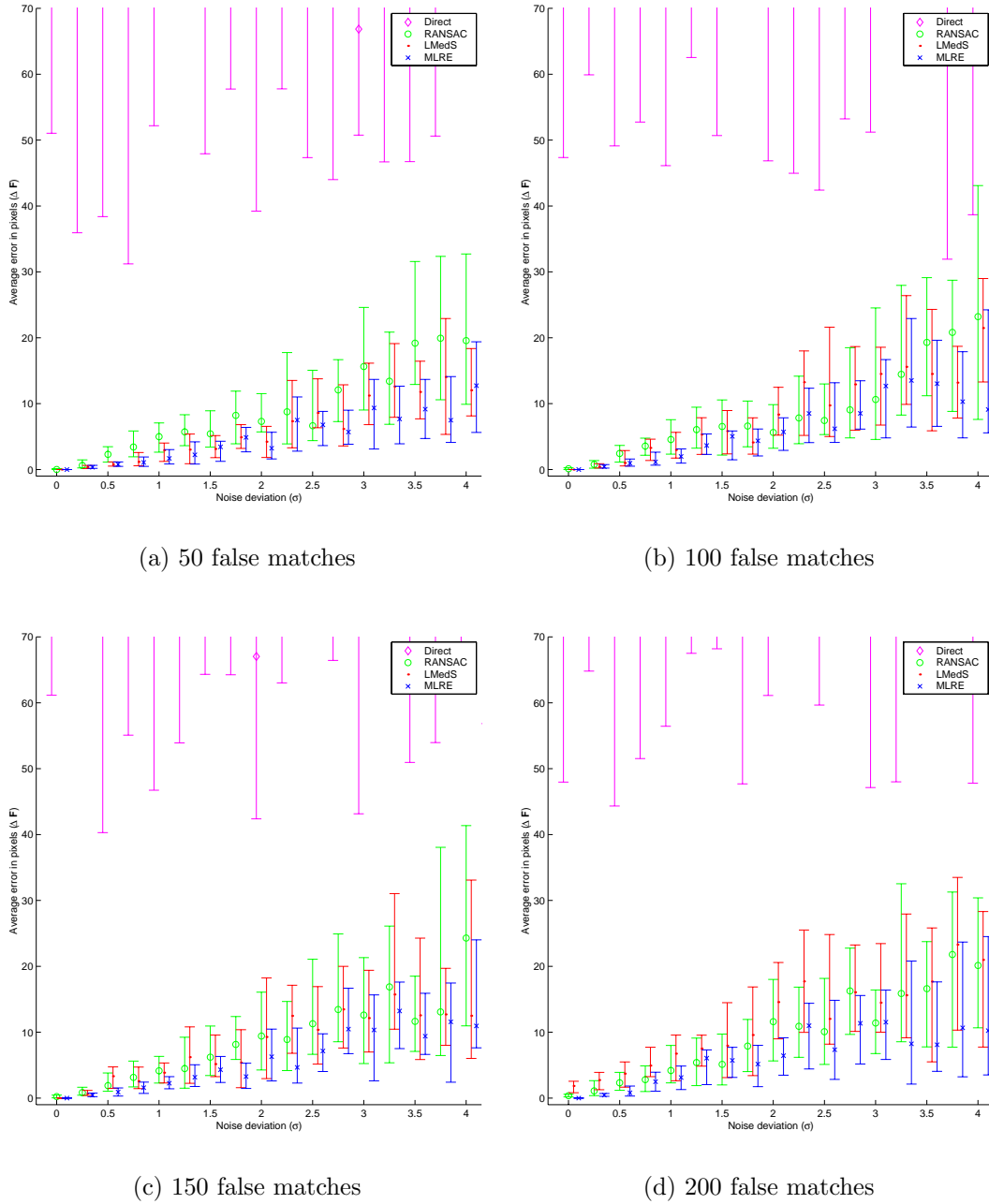
The synthetic data was generated so that 200 points were randomly generated within a cube. The two image planes were simulated by projecting the points using an orthographic projection onto two planes in space that would approximately correspond to a typical stereo view of an object, i.e., the image planes were positioned so that their normal directions deviated from each other by a couple of degrees. To the 2D coordinates of the projections i.i.d. Gaussian noise was added and false matches were randomly generated so that their distribution on the image planes was uniform.

The fundamental matrix estimates were computed in four different ways: directly from (3.4) and (3.5), with the affine RANSAC, affine LMedS (followed by an M-estimator, see Xu and Zhang 1996), and the maximum likelihood robust estimator, where the estimate from the LMedS method was used as the initial guess. RANSAC was implemented as recommended in Hartley and Zisserman (2000), where the number of sampling rounds is determined adaptively and the inlier threshold is set to 1.25 pixels.

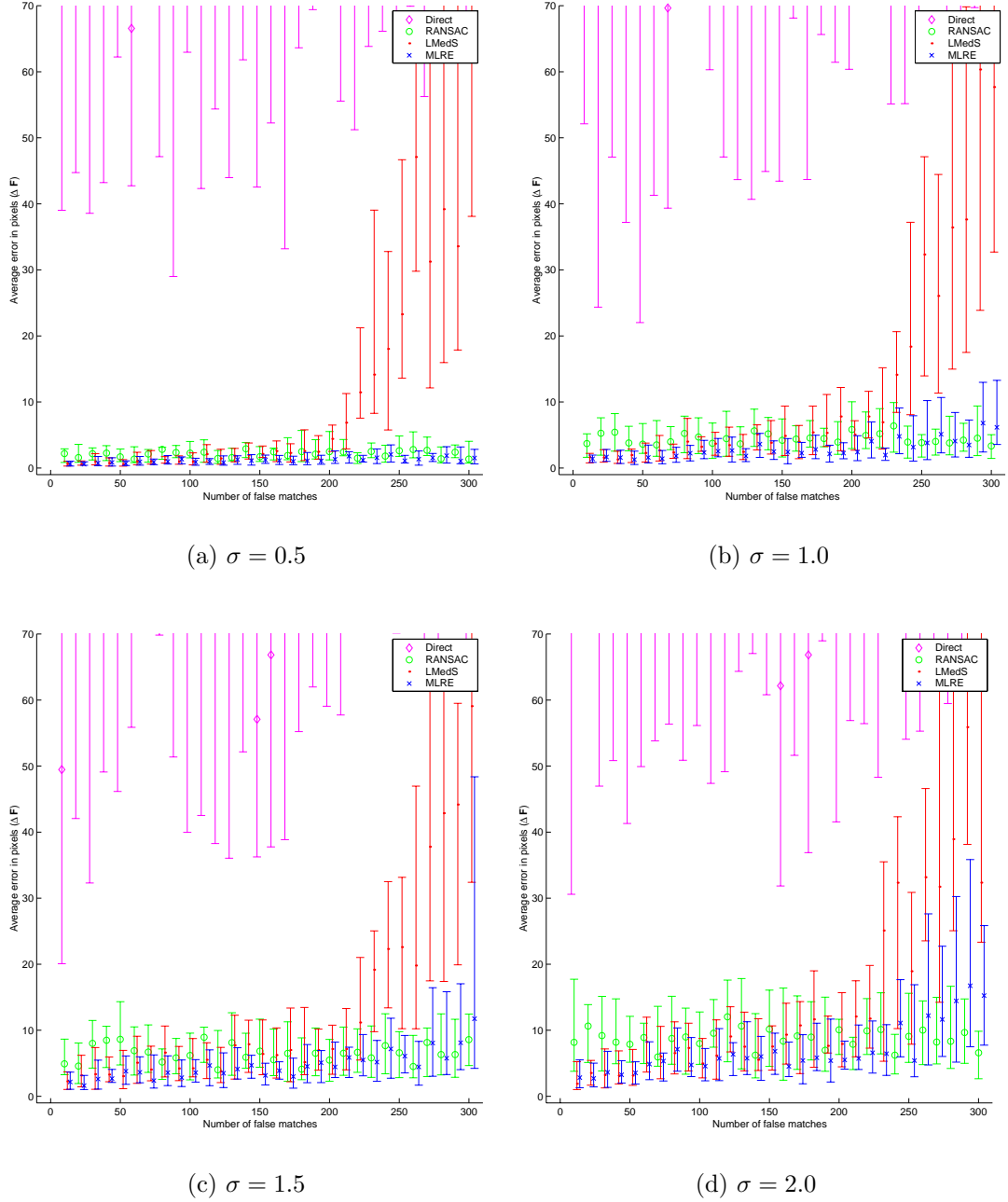
We made two test sets. In the first set, we generated the data with 20 different noise variance values 30 times each and set the number of false matches to 50, 100, 150, and 200. The resulting graphs are shown in Fig. 3.2. The second set was similarly generated but now the number of outliers was varied, and the test was repeated at four different noise levels. The results are shown in Fig. 3.3. In both cases, we used the median of the 30 times repeated experiment as the result whereas the error bounds are the first and third quartile of the observations, respectively.

---

<sup>5</sup>Available from <http://www-sop.inria.fr/robotvis/personnel/zzhang/zzhang-eng.html>.



**Figure 3.2:** Results for the synthetic data where the robustness of the F-matrix is evaluated with varying amounts of noise; deviation of one unit on the  $x$ -axis roughly corresponds to one pixel in a  $500 \times 500$  image. Consequently, the median of 30 times computed error is used whereas the error bounds are the corresponding first and third quartile. The four sub graphs show the results for four different quantities of false matches whereas the number of correct matches is fixed to 200.



**Figure 3.3:** Affine F-matrix estimation with synthetic data. The error is computed for varying quantities of false matches whereas the number of relevant matches is fixed to 200. The median of 30 times computed error is used as a result, and the error bars respectively correspond to the first and third quartile of the observations. In the four sub graphs, the Gaussian noise on the point coordinates are varied from  $\sigma = 0.5$  to 2.0.

It can be seen that the direct method is completely confused by the false matches in the data. Both LMedS and MLRE perform robustly under increasing noise level while the latter is able to produce slightly better results on average. Furthermore, the variance of LMedS is greater than that of MLRE, particularly as the number of false matches grows; the deviation of RANSAC is largest. The affine LMedS works quite robustly under increasing Gaussian noise in contrast to the LMedS with the full camera model (Xu and Zhang 1996, Brandt and Heikkonen 2000b) as long as the number of false matches is kept under 50%. On the other hand, the estimation error grows fast if the number of false matches exceeds 200, as shown in Fig. 3.3. In addition, RANSAC began to experience serious computational problems when the noise deviation exceeded two pixels because of the adaptive sample size determination. At worst, the minimal sets had to be drawn approximately 50 000 times to compute a single F-matrix estimate.

The drawback of the EM implementation of MLRE is that convergence to the global minimum is not guaranteed; however, it is known that the EM solution does not lessen the likelihood of a solution during the iteration. However, local minima represent a problem in all the iterative nonlinear methods that minimize the reprojection error or its approximations in the full projective case. MLRE usually works relatively well, even with poor initial estimates, but to avoid possible local minima and to get the best possible results, the initial estimate should be as good as possible. Therefore, LMedS was originally chosen as the initial estimate with the synthetic data.

The effect of the initial estimate on MLRE can be seen in Fig. 3.3. If the noise level is small, it does not matter which initial estimate is used (Fig. 3.3a) since the results are unchangeable while the initial estimate (LMedS) starts to deteriorate when the number of false matches exceeds 200. If the noise level is large (Fig. 3.3c and 3.3d), the local minima are more probable. This can be seen as there is a notable change in variance of MLRE as the initial estimate deteriorates. The median value is, however, almost unchanged.

On the whole, the results indicate that the maximum likelihood robust estimator is better able to adapt to the data than the affine LMedS or RANSAC. MLRE is assumed to perform even better in practice, because of the slightly pessimistic outlier distribution, as explained above.

### 3.4.2 Real Data

From the INRIA Syntim Database (1999), we chose four image pairs for which the affine F-matrices were estimated with the direct, RANSAC, LMedS, and MLRE methods as with synthetic data. The initial guesses for MLRE were computed by the direct method. They have calibrated stereo rigs so we could use the F-matrices computed from the calibrated data as the ground truth. Moreover, by

**Table 3.1:** Average error between the estimated and calibrated F-matrix.

Method	$\Delta\mathbf{F}$	Method	$\Delta\mathbf{F}$	Method	$\Delta\mathbf{F}$	Method	$\Delta\mathbf{F}$
Direct	35.0	Direct	20.6	Direct	191.1	Direct	71.1
RANSAC	23.0	RANSAC	10.2	RANSAC	10.1	RANSAC	46.8
LMedS	18.1	LMedS	10.3	LMedS	11.8	LMedS	27.9
MLRE	3.9	MLRE	10.6	MLRE	2.5	MLRE	7.1
(a) House		(b) Room		(c) Statue		(d) Array	

using the proposed covariance matrix estimate from Section 3.3, we are better able to visualize the reliability of MLRE by plotting the error bounds for epipolar lines (see Csurka et al. 1997) and observing how the ground truth epipolar lines are related.

From the calibrated outdoor image pair (Fig. 3.4a), 263 point matches were found with the correlation and relaxation techniques (Zhang, Deriche, Faugeras and Luong 1994). The difference between the estimated F-matrices and ground truths are displayed in Table 3.1a. The quantitative ordering of the methods with this stereo pair is the same as it is with the synthetic data. In the relaxation process, some false matches have been neglected so that even the direct method gives a better estimate. Of the four techniques, MLRE still yields the best results with respect to the calibrated fundamental matrix. As expected, the Gaussian residual model for mismatches seem to work better with real images.

In order to visualize the results, epipolar lines were plotted in Fig. 3.4. In Fig. 3.4a, reference points have been marked, and the corresponding epipolar lines are plotted to the other images corresponding to the three estimated F-matrices. It can be seen that the epipolar lines of the direct, RANSAC, and LMedS methods are further from the ground truth than the lines of MLRE. It may be seen that even the 95% error bounds are somewhat narrow and the calibrated lines are inside the epipolar envelopes. The calibrated epipolar lines are also almost parallel, which indicates that the true F-matrix is close to the affine form.

The second test set is shown in Fig. 3.5. Here, 316 point matches were found by correlation and relaxation techniques. If the selected points in the image corners are examined more closely, one can see that the calibrated lines do not intersect the selected points in the other image but go much further away than the epipolar lines computed with the RANSAC, LMedS, and MLRE. This indicates that the calibrated result is not accurate at the image corners and/or the pinhole camera model is not adequate there. In fact, it is likely that the calibration object is put in the center of the view, which makes the results at the image corners uncertain. One should not, therefore, be alarmed by the fact that the calibrated epipolar lines do not always lie inside the epipolar bands, or by the results of Table 3.1b



(a)



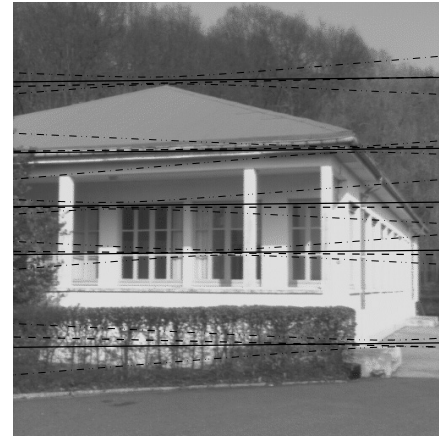
(b) Direct



(c) RANSAC



(d) LMedS



(e) MLRE

**Figure 3.4:** (a) Five points are selected and marked with “+” on one image of a stereo image pair. The corresponding epipolar lines (solid line) in the other image obtained with (b) the direct affine method; (c) affine RANSAC; and (d) affine LMedS. (e) The epipolar lines and the corresponding 95% confidence intervals (dash-dot line) computed with MLRE. The ground truth epipolar lines computed from the calibrated F-matrix are represented by the dashed lines. Image copyrights belong to INRIA-Syntim.



(a)



(b) Direct



(c) RANSAC

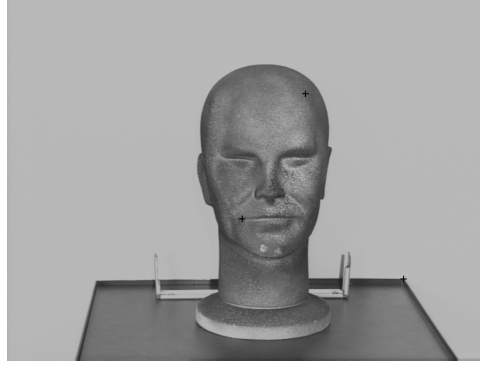


(d) LMedS

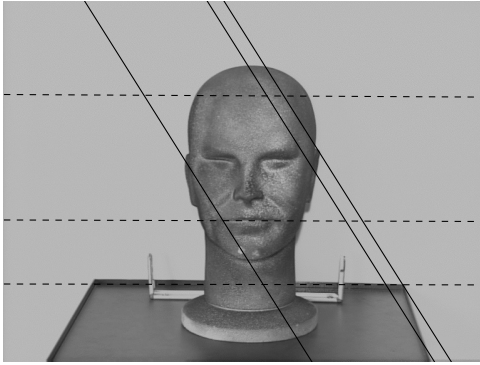


(e) MLRE

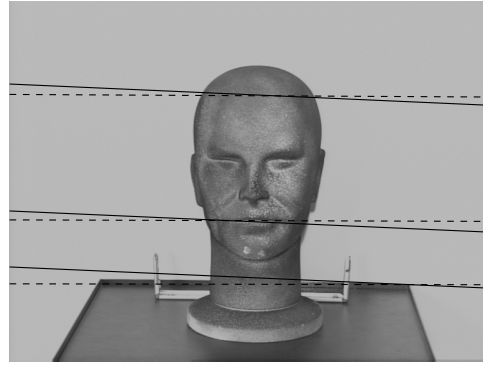
**Figure 3.5:** (a) Five points are marked with “+”. The corresponding epipolar lines (solid line) computed with (b) the direct affine method; (c) affine RANSAC; and (d) affine LMedS. (e) Epipolar lines and their 95% confidence intervals (dash-dot line) computed with MLRE. In the image center, the affine approximation seems to work well but the calibrated epipolar lines (dashed) do not intersect the selected points in the image corners, which indicates inaccuracy of the calibration there. Image copyrights belong to INRIA-Syntim.



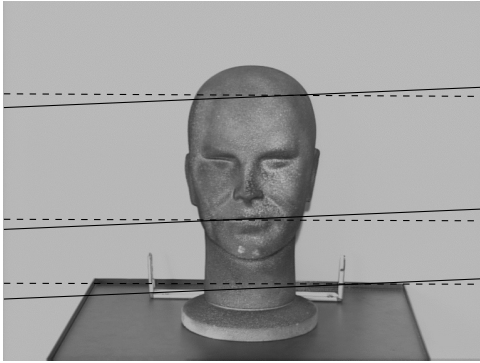
(a)



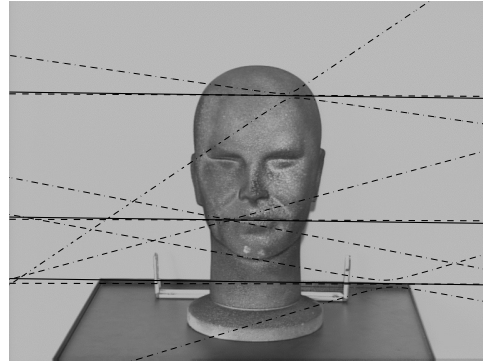
(b) Direct



(c) RANSAC



(d) LMedS



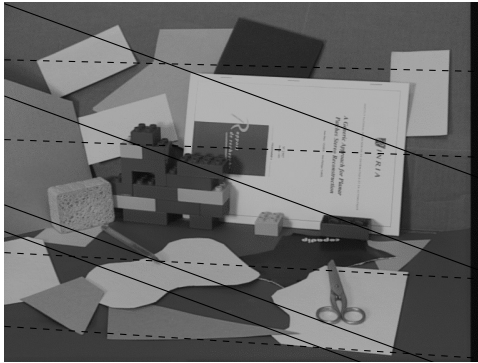
(e) MLRE

**Figure 3.6:** (a) Three points marked with “+”. The other image where the epipolar lines (solid line) have been computed by (b) the direct affine method; (c) affine RANSAC; and (d) affine LMedS. (e) Epipolar lines and their 70% confidence intervals (dash-dot line) computed by MLRE. The confidence intervals are comparatively large because there are only small variations in the depth. The calibrated epipolar lines are represented by the dashed lines. Image copyrights belong to INRIA-Syntim.

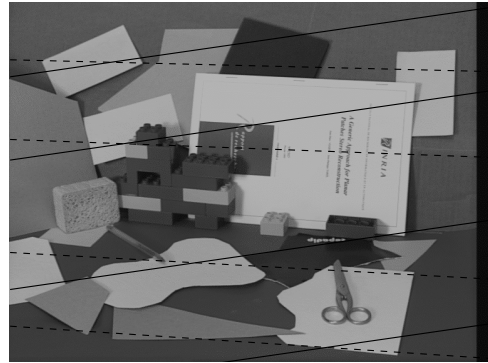




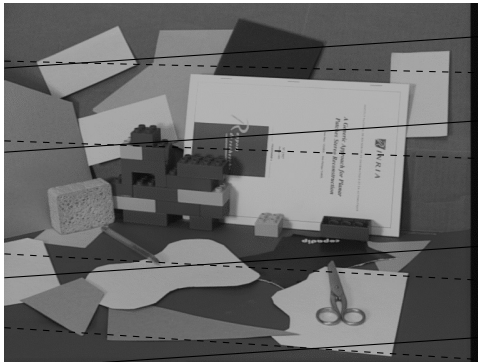
(a)



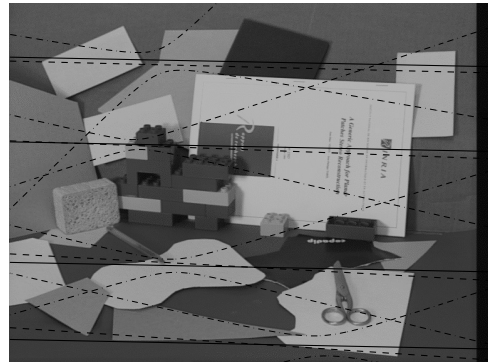
(b) Direct



(c) RANSAC



(d) LMedS



(e) MLRE

**Figure 3.7:** (a) Four points are marked with “+”. The corresponding epipolar lines (solid line) computed by (b) the direct affine method; (c) affine RANSAC; and (d) affine LMedS. (e) Epipolar lines and their 70% confidence intervals (dash-dot line) computed by MLRE. The calibrated epipolar lines are dashed. Image copyrights belong to INRIA-Syntim.

because in the middle region, the epipolar lines of the robust methods match well with the calibrated lines.

The third stereo image pair is taken of a statue (Fig. 3.6). Now, we obtained 195 matching point pairs with the correlation and relaxation techniques. The results in Table 3.1c indicate that the direct method failed to give a reasonable estimate totally. Even though the results obtained by RANSAC and LMedS are tolerable, MLRE showed even better performance. The same can be observed in Fig. 3.6. The affine approximation is good here because the epipolar lines are almost parallel and the epipolar lines computed by the proposed method get closest to the ground truth. However, the error bounds are somewhat large here although we used only 70% confidence intervals. This is likely to be due to the small variation in the depth.

The last image pair representing an array is displayed in Fig. 3.7. With this image pair, the correlation and relaxation techniques gave the fewest point pairs of the four stereo sets, namely 132. The results (Table 3.1d) again indicate that the proposed method is closest to the calibrated F-matrix. The same can be found in Fig. 3.7, where the epipolar lines are plotted. The confidence intervals are moderate in this case and the calibrated result is well inside the epipolar band like before.

### 3.5 Projective Experiments

In this section, we consider fundamental matrix estimation using the projective camera model. Since the normal distribution is not theoretically (Brandt and Heikkonen 2000b) as good an approximation of the outlier distribution as it is in the affine case, we model the outlier distribution with a general Gaussian mixture as proposed in the previous chapter. The F-matrices were therefore estimated using the Algorithm 2 from corner matches, obtained by correlation and relaxation (Zhang et al. 1994) as in the affine case.

To keep the procedure computationally simple, we consider here the first order (Sampson) approximation of the reprojection error. The residual is then

$$\epsilon = \frac{\mathbf{m}'^T \mathbf{F} \mathbf{m}}{(l_1^2 + l_2^2 + l_1'^2 + l_2'^2)^{1/2}}, \quad (3.32)$$

where  $(l_1 \ l_2 \ l_3)^T = \mathbf{F} \mathbf{m}$  and  $(l_1' \ l_2' \ l_3')^T = \mathbf{F}^T \mathbf{m}'$ .

Since no ground truth was available, the matches were randomly divided into independent train (90%) and test sets (10%) where the train sets were only used in the estimation. The test sets were manually cleaned from mismatches and used only in the evaluation. To achieve a feasible estimate for the F-matrix, we used the minimal parameterization (Zhang and Loop 2001) in the optimization. The

**Table 3.2:** Results for the fundamental matrix estimation with real data. Error is the standard deviation of the correct residuals in the independent test set.

Method	Error	Method	Error
LMedS	0.61	LMedS	0.61
MLRE	0.54	MLRE	0.61

(a) Footpath

(b) Cathedral tower

**Table 3.3:** Test set error of the maximum likelihood robust estimator and evidence (in the train set) with residual models of varying complexity. The MDL selections are bolded.

Kernels	Error	Evidence	Kernels	Error	Evidence
1	6.5	335.4	1	6.6	338.7
2	0.54	210.6	2	0.61	199.4
3	0.54	208.7	<b>3</b>	<b>0.61</b>	<b>196.7</b>
<b>4</b>	<b>0.54</b>	<b>208.7</b>	4	0.61	196.9
5	0.54	209.3	5	0.61	197.5

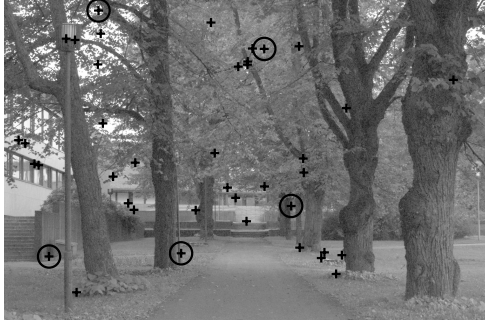
(a) Footpath

(b) Cathedral tower

initial estimate of model parameters was computed with the LMedS estimator and the number of kernels selected in the residual mixture model was from one to five. In the experiment, we used two pairs of real images shown in Fig. 3.8 and 3.9. The proposed estimator was compared to the LMedS followed by an M-estimator (see Xu and Zhang 1996). Consequently, we used the standard deviation of the residuals in the test set. The results are in Table 3.2.

It can be seen that MLRE yielded good results in both cases; not surprisingly, LMedS similarly works well. In Table 3.3, we have the scores for different quantities of kernels in the residual models. In these cases, the number of kernels in the mixture model is not a very strict parameter since the error is almost the same with all selections from two to five. The evidence, penalized negative log likelihood, is therefore also somewhat flat.

As with affine cameras, the covariance estimate for the fundamental matrix can be computed with MLRE without having to *erroneously* classify points as good and false. This can be done in a similar way to in (Csurka et al. 1997) simply by replacing the cost function  $\sum_i C_i^2$  (on page 23) with the posterior weighted function  $\sum_i P(\mathcal{R}|\epsilon_i, \omega') C_i^2$ . Confidence intervals for the epipolar lines can thereafter be computed (Csurka et al. 1997), as illustrated in Fig. 3.8b and 3.9b.



(a)



(b)

**Figure 3.8:** (a) Left image of the footpath pair with the points in the cleaned *test set* plotted. Randomly selected points, whose epipolar lines are plotted in the second image, are circled. (b) Five estimated epipolar lines and their 95% confidence intervals. The point near the epipole has a larger uncertainty than expected.



(a)



(b)

**Figure 3.9:** (a) Left image of the cathedral tower pair and the test set points plotted. (b) Epipolar lines with the 95% confidence intervals.

## 3.6 Summary and Discussion

In this chapter, we have evaluated the maximum likelihood robust estimator in estimating the geometry of two views, represented by the fundamental matrix. Both affine and projective camera models were considered; they conveniently represent the cases of known and unknown residual models, respectively, as discussed in the previous chapter. As expected, after the proof of optimality, the affine experiments with both synthetic and real data reveal that MLRE gave better results than LMedS and RANSAC, which are perhaps the most considered robust estimators for image geometry estimation so far.

In the projective experiments, the outlier distributions were somewhat flat, and even a two kernel mixture would have been a good selection for the residual model. The MDL principle selected a slightly more complex residual model for both cases but the result was practically the same, perhaps because, as far as these results are concerned, it is not very important if some distant hills in the outlier distribution are modeled more accurately by using more kernels or not. In spite of our preliminary pessimistic discussion in Brandt and Heikkonen (2000b), where we rushed by claiming that two kernel model is too simple for the residual of the projective F-matrix in theory, after all, it seems to be a suitable model in practice. Experiments by Torr and Murray (1997) supports this too.

Let us summarize the advantages, based on the optimality, of the maximum likelihood robust estimator in the F-matrix estimation: (1) it has an *adaptive* mechanism to model the residual of false matches, where (2) no assumption of symmetry for the total residual distribution is made. (3) It does not loosen the normality assumption for the relevant matches because individual matches are not classified as relevant or false. Moreover, (4) the covariance matrix of the F-matrix (used e.g. in self-calibration, projective reconstruction, and computing confidence intervals for epipolar lines, see Csurka et al. 1997) can be approximated by generalizing the uncertainty computations of Shapiro et al. (1994) and Shapiro and Brady (1995) (affine camera), and Csurka et al. (1997) (projective camera), as was proposed in the chapter. The covariance matrix estimation is not *directly* possible in those methods which use random sampling or M-estimators (Xu and Zhang 1996, Torr and Murray 1997). In these cases, the matches must be first classified as relevant or false, and thereafter the normal distribution assumption is used for the relevant classified matches in estimating the covariance, nevertheless the assumption does not strictly hold.

In general, to draw conclusions as to how well the proposed estimator really works in practice with respect to, for example, the LMedS estimator, much more research would be needed. The results of this chapter are promising since LMedS is known to be very good and robust because of the median and the following M-estimator. The maximum likelihood robust estimator is most efficient when the

form of outlier residual model is known and a good initial estimate is available. The more general implementation of the maximum likelihood robust estimator must consider both the density estimation problem as well as model selection. One way of solving these problems, possibly not the best, was also proposed in the previous chapter. Anyhow, because the proposed maximum likelihood robust regression was proven optimal under certain conditions, it forms a universal principle for regression from fatally contaminated data.

## Chapter 4

# Uncertainty of the F-Matrix and Disparity of the Scene<sup>1</sup>

After proposing in the previous chapter how the covariance matrix of the fundamental matrix should be automatically computed from outlier and noise contaminated data, in this chapter we will focus on the utilization of the uncertainty information. We first show in Section 4.2 that the eigenvectors of the epipolar line covariance matrix have certain interpretations in the other image that reflects the uncertainty, and thereby the disparity, of the points that have been used in the epipolar geometry estimation. We show that the location of the image origin plays an unwanted special role, and propose how this can be released in Section 4.3. We finally suggest (Section 4.4) how the uncertainty of the epipolar geometry could be used in matching, and conclude the chapter by deriving the probability density from the epipolar line covariance that explicitly represents a weak point–point constraint between two views.

### 4.1 Introduction

Epipolar geometry helps in matching points between two views since the correspondence for a point in the first view must lie on the corresponding epipolar line in the second view. However, when the epipolar geometry is estimated from noisy data, not even the noise free matching points will lie on the corresponding epipolar line precisely, since the fundamental matrix will be known only up to a finite level of accuracy. The covariance matrix of the F-matrix characterizes this uncertainty and it is possible to compute error bounds for the epipolar lines from the covariance (Csurka et al. 1997, Hartley and Zisserman 2000).

---

<sup>1</sup>This chapter is the most recent of this thesis and the results are previously unpublished. However, the first ideas are outlined in Brandt and Heikkonen (2001a).

The bounds for the epipolar lines or the epipolar envelope usually represents a hyperbola, and it has been noticed by Xu and Zhang (1996), Zhang (1998a), and Faugeras and Luong (2001) that the epipolar envelope seem to be narrowest at the most probable location of the match. They suggest that the covariance matrix of the fundamental matrix might therefore capture information on the disparity of the scene, although the epipolar geometry itself reveals nothing about the disparity. However, an accurate explanation for the narrowest point of the envelope has not yet been given (Hartley and Zisserman 2000). As we will see in the following sections, the conjecture is indeed correct, and we will show how the most probable location can be computed.

In effect, our intuition can explain the previous observation. As we compute the epipolar lines in the one view, corresponding to a point in the other view, it reflects a common line fitting situation: the error bounds are small and narrow where we have had observations, i.e., the interpolation error would be small but the extrapolation error large. As another example, we will observe in Chapter 6 that, with the affine camera model, the ML estimates for the inhomogeneous projection matrices can be understood as a product of the sample cross-covariance matrix between the measured image coordinates and 3D coordinates, and the sample covariance matrix of the 3D reconstructions. Because the projection matrices of two views determine the fundamental matrix uniquely, there must also be an analytical relationship between the covariance of the fundamental matrix and covariance of the reconstructions of the points used in the estimation.

The theoretical analysis of the epipolar line covariance, derived from the fundamental matrix covariance, will be provided in the following section. It turns out that, in effect, each eigenvector of the epipolar line covariance matrix has *two* interpretations in the image, a point and a line interpretation. One of these interpretations is the narrowest point of the epipolar envelope. However, the location of the image origin affects these interpretations as well as the epipolar envelopes, which represents an undesirable effect; hence, the correct fixation of the origin will be discussed in Section 4.3. After showing that the fundamental matrix covariance does in fact contain disparity information, in Section 4.4, we suggest how it could be utilized in matching. By using a convenient Gaussian approximation in the dual space, we derive a probability distribution for the correspondence point location in the other view.

Recently, Triggs (2001) proposed a method that models the joint probability distributions of correspondences. As he states, they aim to summarize the observed behavior of the given training correspondences, but not to rigidly constrain them to an ideal predefined geometry; this also applies exactly to our approach. In both approaches, a loose analogy can be seen to Bayesian inference (see e.g. Bishop 1998) in the form of model averaging (Torr and Zisserman 1998) and marginalization. The difference between our and Triggs's (2001) approach is significant since he models the feature distributions by Gaussians on the image planes, and



considers the algebraic linear system for their estimation. We, however, derive the corresponding distributions from the estimated covariance matrix in the dual space. The advantage of our approach is that we may use the statistically correct criterion for the fundamental matrix and its covariance, and are able to estimate them robustly, as proposed in the previous chapter.

## 4.2 Special Points and Lines

To create a convenient parameterization for lines, we follow Hartley and Zisserman (2000) by first defining the point–line relation of the epipolar geometry using the mapping  $\mathbf{l} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  such that  $\mathbf{l} = \mathbf{l}(\mathbf{m}) = \|\mathbf{F}\mathbf{m}\|^{-1}\mathbf{F}\mathbf{m}$ . Epipolar lines may hence be uniquely represented by points on the unit sphere, and the lines through the image origin do not need special parameterization, as in Csurka et al. (1997) and our preliminary considerations (Brandt and Heikkonen 2001a).

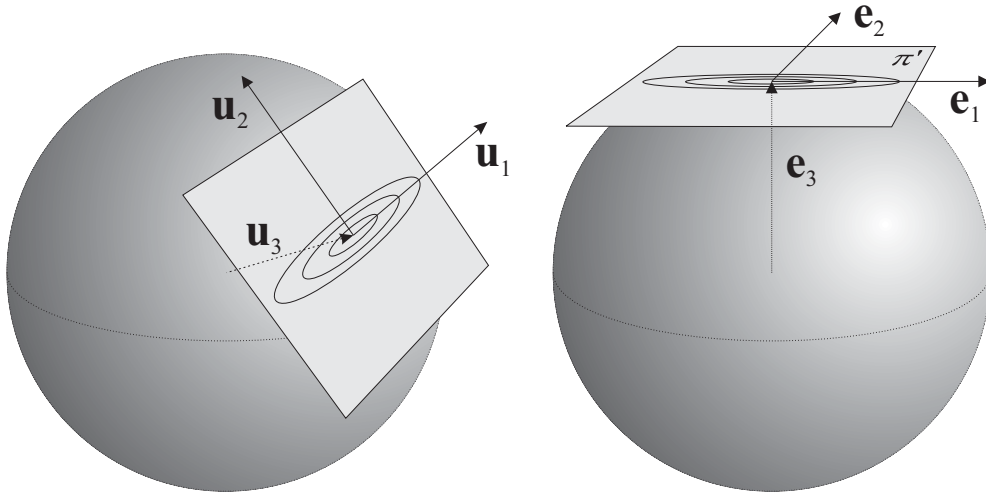
The first-order covariance approximation for the epipolar line  $\mathbf{l}$  in the second image, given the point  $\mathbf{m}$  in the first, is given by

$$\mathbf{C}_1 \simeq \frac{\partial \mathbf{l}}{\partial \mathbf{F}} \mathbf{C}_F \frac{\partial \mathbf{l}}{\partial \mathbf{F}}^T + \frac{\partial \mathbf{l}}{\partial \mathbf{m}} \mathbf{C}_m \frac{\partial \mathbf{l}}{\partial \mathbf{m}}^T, \quad (4.1)$$

where it is assumed that  $\mathbf{m}$  is contaminated by Gaussian noise with covariance matrix  $\mathbf{C}_m$ . In the equation,  $\partial \mathbf{l} / \partial \mathbf{F}$  stands for the Jacobian of the point–line mapping above with respect to the  $\mathbf{F}$ , which is here regarded as a vector of 9 elements.

**Proposition 4.2.1** *Let  $\mathbf{m}$  be a point in the first view, corrupted by Gaussian noise with covariance matrix  $\mathbf{C}_m$ . Let  $\hat{\mathbf{F}}$  be the estimated fundamental matrix with estimated covariance matrix  $\hat{\mathbf{C}}_F$ , where the sought correspondence  $\mathbf{m} \leftrightarrow \mathbf{m}'$  has a priori the same (disparity) distribution as the point matches with which  $\hat{\mathbf{F}}$  and  $\hat{\mathbf{C}}_F$  have been estimated. Then, up to the first-order and Gaussian approximation, the eigenvectors, sorted in descending order, of the epipolar line covariance matrix  $\hat{\mathbf{C}}_1$ , corresponding to  $\mathbf{m}$  and having the rank of 2 with no multiple eigenvalues, have the following interpretation in the second image:*

- $\mathbf{u}_1$  : *the least probable point match for  $\mathbf{m}$  on the estimated epipolar line;  
the least probable epipolar line that intersects the point  $\mathbf{u}_2$*
- $\mathbf{u}_2$  : *the most probable point match for  $\mathbf{m}$  in the second image;  
the least probable epipolar line that intersects the point  $\mathbf{u}_1$*
- $\mathbf{u}_3$  : *the least probable pencil of epipolar lines;  
the estimated (mean) epipolar line corresponding to  $\mathbf{m}$ .*



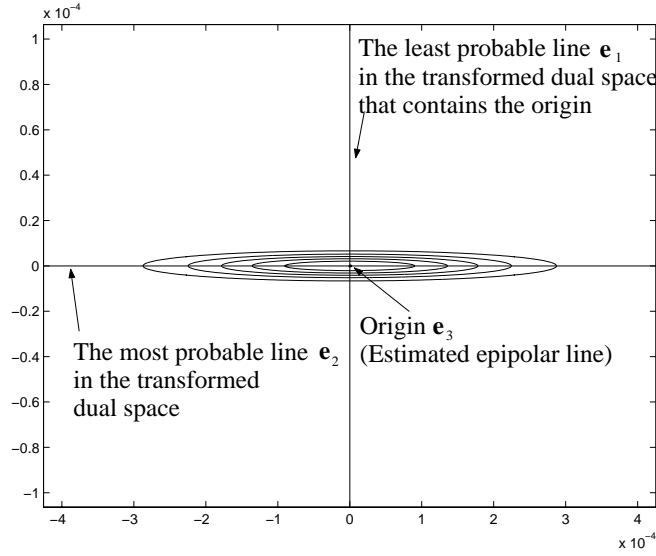
**Figure 4.1:** Epipolar line parameterization with the unit sphere in the dual space (left) and the transformed dual space. The tangent plane at the estimated epipolar line  $\mathbf{l}_0 \equiv \mathbf{u}_3$  is used in the first order approximation of the epipolar line covariance where the axes  $\mathbf{u}_1$  and  $\mathbf{u}_2$  indicate the directions of the largest and smaller variance, respectively. We get into the transformed dual space (right) by  $\mathbf{U}^T$ .

*Proof.* Since the first-order approximation for  $\mathbf{C}_1$  is equivalent to its tangential approximation at the estimated (mean) epipolar line  $\mathbf{l}_0 = \|\mathbf{F}\mathbf{m}\|^{-1}\mathbf{F}\mathbf{m}$ , the covariance matrix approximation is degenerate and positive semi definite. Since it is also real and symmetrical, we may perform the similarity transform and define

$$\mathbf{C}'_1 = \mathbf{U}^T \mathbf{C}_1 \mathbf{U} = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (4.2)$$

where  $\mathbf{U}$  contains the eigenvectors of  $\mathbf{C}_1$  normalized to the unit norm and ordered such that  $\sigma_1 \geq \sigma_2$ . We notice that the mean epipolar line  $\mathbf{l}_0$  belongs to the left null space of the Jacobians  $\partial \mathbf{l} / \partial \mathbf{F}$  and  $\partial \mathbf{l} / \partial \mathbf{m}$ ; so it must also belong to the null space of  $\mathbf{C}_1$  (see Fig. 4.1), hence,  $\mathbf{u}_3 \equiv \mathbf{l}_0$ . Let the tangent plane of the unit sphere at  $\mathbf{u}_3$  be  $\pi$  and the tangent plane at  $\mathbf{e}_3 = (0 \ 0 \ 1)^T$  be  $\pi'$ . If we regard  $\mathbb{P}^2$  as a projective space  $\mathbb{P}^2$ ,  $\mathbf{U}^T$  can be also seen as the collineation  $\mathbf{U}^T : \mathbb{P}^2 \longrightarrow \mathbb{P}^2$  from the projective plane  $\pi$  to  $\pi'$ , i.e., the lines (and points) transform according to  $\mathbf{l}' = \mathbf{U}^T \mathbf{l}$ . The mean epipolar line is the origin in the transformed dual space  $\pi'$  since  $\mathbf{e}_3 = \mathbf{U}^T \mathbf{u}_3$ . Correspondingly, the direction of the largest variance  $\mathbf{u}_1$  is transformed to  $\mathbf{e}_1 = (1 \ 0 \ 0)^T$  and the direction of the smaller variance  $\mathbf{u}_2$  to  $\mathbf{e}_2 = (0 \ 1 \ 0)^T$ .

It is assumed that the covariance is non-isotropic in the sense that  $\sigma_1 > \sigma_2$ . Therefore, in the transformed dual space the most probable line is the  $x$ -axis  $(0 \ 1 \ 0)^T \equiv \mathbf{e}_2$ , since the *marginal probability* or integral of the two-dimensional



**Figure 4.2:** Illustration of the transformed dual space or the projective plane  $\pi'$  with the contours of the Gaussian corresponding to  $\mathbf{C}'_1$ . The estimated epipolar line  $\mathbf{l}_0$  corresponds to the origin  $\mathbf{e}_3$ , and the most probable point in the image space is  $\mathbf{u}_2 = \mathbf{U}\mathbf{e}_2$  since the marginal probability over the line  $\mathbf{e}_2$  is largest. Correspondingly, the least probable point on the estimated epipolar line is  $\mathbf{u}_1 = \mathbf{U}\mathbf{e}_1$  since  $\mathbf{e}_1$  has the smallest total probability of the lines that contain the origin.

(degenerate three-dimensional) Gaussian distribution with the covariance matrix  $\mathbf{C}'_1$  centered at  $\mathbf{e}_3$  is the largest over this line (see Fig. 4.2). The most probable point in the image hence is  $\mathbf{u}_2 = \mathbf{U}\mathbf{e}_2$  in the sense that the *total probability of the pencil of epipolar lines* intersecting  $\mathbf{u}_2$  is the largest. Since the (dual) line  $\mathbf{e}_2$  intersects the origin, it contains the (dual) point of the mean epipolar line, and the point  $\mathbf{u}_2$  is therefore also on the estimated epipolar line. The least probable point on the estimated epipolar line is correspondingly the  $y$ -axis or the line  $\mathbf{e}_1$  in the transformed dual space since the marginal probability is the smallest of all (dual) lines that intersect the origin  $\mathbf{e}_3$ . Hence, the least probable pencil of epipolar lines that contain the line  $\mathbf{u}_3$  must be  $\mathbf{u}_1 = \mathbf{U}\mathbf{e}_1$ . The least probable line in the transformed dual space is the line at infinity  $\mathbf{e}_3$ , its marginal probability being zero; hence the *point interpretation* of  $\mathbf{u}_3$  is the least probable pencil of epipolar lines in the second image. In the transformed dual space,  $\mathbf{e}_1$  is the least probable point on the  $x$ -axis because, as an ideal point it has zero probability density value. Its line interpretation  $\mathbf{u}_1$  in the second image is hence the least probable line that intersects the most probable point  $\mathbf{u}_2$ . Correspondingly,  $\mathbf{e}_2$  is the least probable point on the  $y$ -axis in the dual space, and therefore  $\mathbf{u}_2$  is the least probable epipolar line for  $\mathbf{m}$  that intersects the point  $\mathbf{u}_1$ .

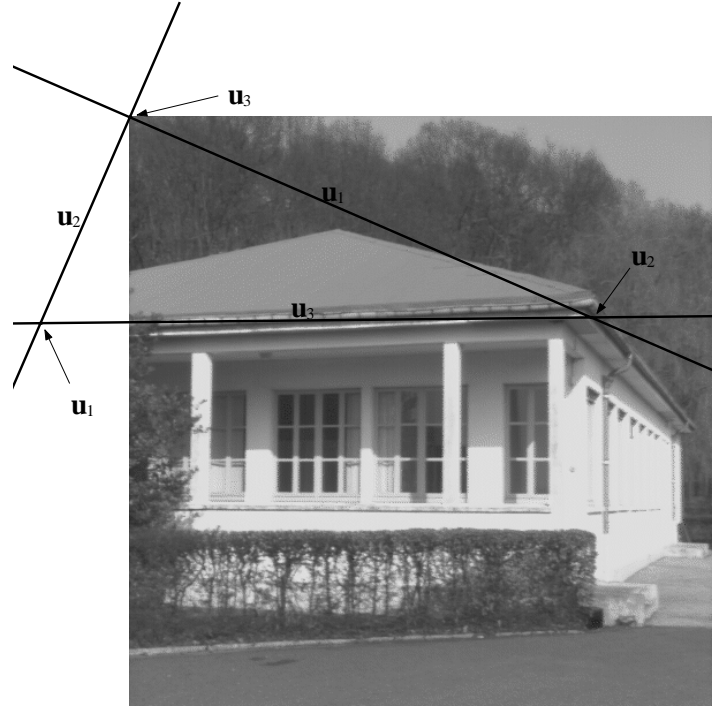
**Table 4.1:** Summary of the relations between the eigenvectors of the epipolar line covariance matrix  $\mathbf{C}_1$  corresponding to the point  $\mathbf{m}$  in the first image. Each interpretation can be obtained by taking cross products between the other two dual interpretations. For example, the least probable pencil of epipolar lines  $\mathbf{u}_1$  that contains the estimated epipolar line (point interpretation) is obtained by intersecting the estimated epipolar line  $\mathbf{u}_3$  and the least probable epipolar line  $\mathbf{u}_2$  or equivalently joining the origin  $\mathbf{e}_3$  and ideal point  $\mathbf{e}_2$  on the  $y$ -axis in the transformed dual space.

	Interpretation in the image	Interpretation in the transformed dual space
$\mathbf{u}_1 = \mathbf{u}_2 \times \mathbf{u}_3$	The least probable match for $\mathbf{m}$ on the estimated epipolar line; The least probable epipolar line that intersects the point $\mathbf{u}_2$	The $y$ -axis; The ideal point of the $x$ -axis
$\mathbf{u}_2 = \mathbf{u}_3 \times \mathbf{u}_1$	The most probable match for $\mathbf{m}$ in the second image; The least probable epipolar line that intersects the point $\mathbf{u}_1$	The $x$ -axis; The ideal point of the $y$ -axis
$\mathbf{u}_3 = \mathbf{u}_1 \times \mathbf{u}_2$	The least probable pencil of epipolar lines; The estimated (mean) epipolar line corresponding to $\mathbf{m}$ .	The line at infinity; The origin

The relations between the point  $\mathbf{m}$  and the eigenvectors of the corresponding epipolar line covariance matrix  $\mathbf{C}_1$  are conveniently coupled such that each point interpretation of the eigenvectors can be obtained by intersecting the other two line interpretations, and *vice versa*. The relations are summarized in Table 4.1.

A crucial assumption in the proof is that the non-zero eigenvalues of the epipolar line covariance matrix are not equal. If they were equal, no line intersecting the origin in the transformed dual space would be in a more probable position. In other words, every pencil of lines centered at any point on the mean epipolar line would share equal total probability; in this sense the fundamental matrix covariance matrix would reveal nothing about the disparity of the scene. Nevertheless, the fundamental matrix is normally estimated from point matches with concentrated disparity distribution, which suggests that  $\sigma_1 \gg \sigma_2$ , which can be easily verified experimentally.

In Fig. 4.3, we have an example of how the point and line interpretations of the eigenvectors of the epipolar line covariance are related. Since the distance between the origin and the estimated epipolar line is large, the point interpretation of  $\mathbf{u}_3$  is very close to the origin. This example shows that, while the origin may be fixed to an arbitrary point in the image, its location exerts a special influence, and affects the eigenvectors of  $\mathbf{C}_1$  and hence the location of the special points

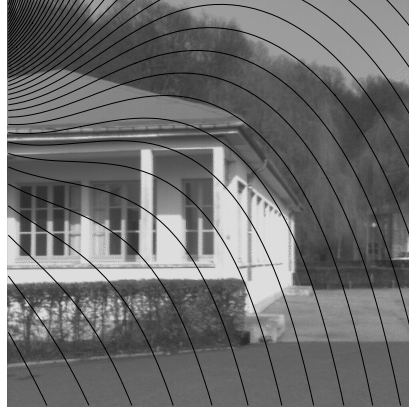


**Figure 4.3:** Point and line interpretations of the eigenvectors of the epipolar line covariance matrix  $\mathbf{C}_1$  that is computed for the point in the cornice of the roof shown in Fig. 3.4a. As expected, the most probable point  $\mathbf{u}_3$  is close to the true correspondence point. Since the estimated epipolar line  $\mathbf{l}_0 \equiv \mathbf{u}_3$  is far away from the origin, the point interpretation of the point  $\mathbf{u}_3$  is correspondingly very close to the origin. This makes the origin a special point of the second image that affects the point and line interpretations of the eigenvectors.

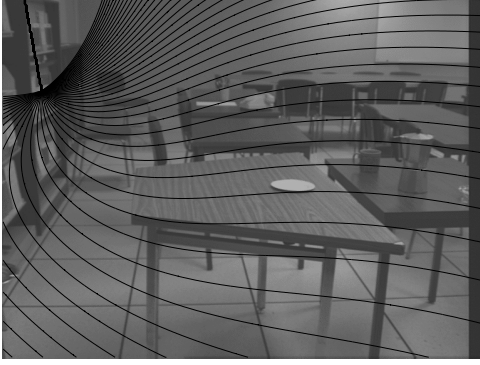
and lines in the image, except  $\mathbf{l}_0$ , and even the epipolar envelopes, which is not desirable at all. How we can overcome this problem is discussed in the following section.

The special role of the origin can be seen from Fig. 4.4, where we have plotted contours of the disparity between points and the corresponding most probable points. We have used the affine approximation in these images (see Section 3.4), hence the linear covariance approximation of the fundamental matrix has at most four linearly independent vectors. The corresponding contours of the projective model should be more flexible to the scene disparity because the projective covariance approximation correspondingly has seven linearly independent vectors.

Let us now return to the previously made hypothesis that the match location is with higher probability at the narrowest point of the epipolar envelope since we are now able explain this observation. The confidence intervals or equi-likelihood



(a)



(b)

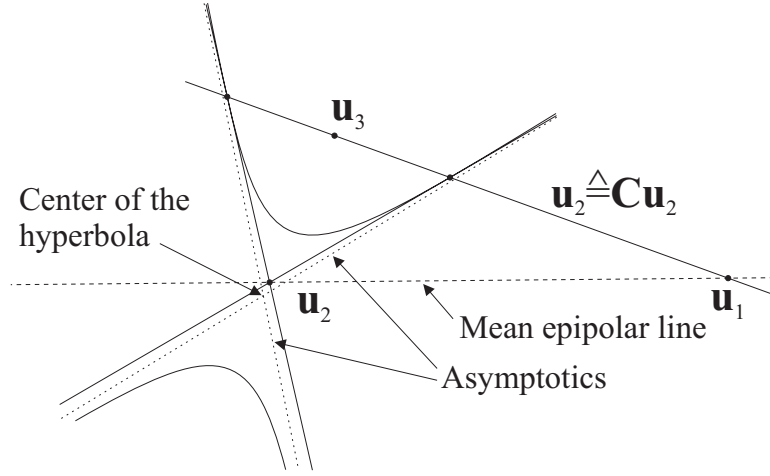


(c)



(d)

**Figure 4.4:** Disparity contours between these images and the most probable point  $\mathbf{u}_2$  in the second image. The disparity has been determined by computing the most probable point for each pixel in the first image and calculating their distance in the rectified coordinate frame. The contours summarize the overall disparity information captured by the covariance matrix of the fundamental matrix, even though here only the maximum ( $\mathbf{u}_2$ ) of the point–point distribution has been used (see Section 4.4). The special role of the image origin is also distinguishable.



**Figure 4.5:** The pole–polar relationship suggests by definition that any point  $\mathbf{m}$  outside the conic  $\mathbf{C}$  induces a polar line  $\mathbf{l} = \mathbf{C}\mathbf{m}$  that intersects the conic at two points, and the tangents at these points intersect at  $\mathbf{m}$ . Since here  $\mathbf{u}_2$  is an eigenvector of the epipolar envelope  $\mathbf{C}$  and is outside the conic, its line interpretation is polar to its point interpretation. Moreover, if  $\mathbf{C}$  is a hyperbola and the point  $\mathbf{u}_3$  is far away from the epipolar line (if necessary, this is achieved by translating the origin sufficiently far away from the epipolar line, see Section 4.3), the tangent lines of  $\mathbf{C}$  at the intersection points of the line  $\mathbf{u}_2$  and  $\mathbf{C}$  are close to the asymptotics of  $\mathbf{C}$ . The center of the hyperbola, the intersection of those tangent lines whose points of tangency are on  $\mathbf{l}_\infty$ , is then close to the point  $\mathbf{u}_2$ .

contours of the epipolar lines in the first order approximation are represented by the conic (Hartley and Zisserman 2000)

$$\mathbf{C} = \mathbf{l}_0 \mathbf{l}_0^T - k^2 \mathbf{C}_1, \quad (4.3)$$

where  $k^2$  follows the cumulative  $\chi_n^2$  distribution with two degrees of freedom. We observe that the line interpretation of  $\mathbf{u}_2$  is polar of the point  $\mathbf{u}_2$  with respect to the conic  $\mathbf{C}$  (see Fig. 4.5) because  $\mathbf{C}$  has identical eigenvectors to  $\mathbf{C}_1$ ; hence,  $\mathbf{u}_2 \hat{=} \mathbf{C}\mathbf{u}_2$  (equality up to scale). Let us additionally assume that the epipolar envelope is a hyperbola of two sheets, as is normally the case. Since the least probable epipolar line is normally far away from the most probable point, the tangency lines from the most probable point are close to asymptotics of the hyperbola. The center of the hyperbola is the intersection of its asymptotics, i.e., the line at infinity is polar to the center. The center of the hyperbola is therefore close to the most probable point, but not exactly, unless the line  $\mathbf{u}_2$  is the line at infinity.

**Corollary 4.2.2** *If  $\mathbf{u}_2$  represents the image origin, or equivalently the line at infinity, the epipolar envelope is always a hyperbola whose center is at the origin.*

*If the epipolar envelope is a hyperbola, its center coincides with the point  $\mathbf{u}_2$  if and only if  $\mathbf{u}_2$  represents the origin.*

### 4.3 How to Fix the Origin of the Image

A direct consequence of using only the first-order approximation of the epipolar line covariance is the special role of the origin in the second image. This is highly undesired since the origin may be set as an arbitrary point in the image. It hence suggests that one should fix the origin to coincide with one of the special points derived in the previous section. Of the three points, at least  $\mathbf{u}_2$  should be a finite point. However, since its accurate location is affected by the location of the origin, it is not quite clear where the origin should be set. Therefore, we choose different tactics.

As we learned in the previous section, the point  $\mathbf{u}_3$  is near the origin if the distance of the mean epipolar line is large. Since the point interpretation of  $\mathbf{u}_3$  is the least probable point in the second image, it should be furthest away from the estimated epipolar line. A convenient way of achieving this is simply to translate the origin away from the image in the normal direction of the mean epipolar line. Such translation implies better validity of the tangential approximation of the epipolar line covariance since the epipolar line variation agglomerates to a smaller neighborhood on the surface of the unit sphere.

We thus define a homogeneous translation matrix

$$\mathbf{T}_n = \begin{pmatrix} 1 & 0 & n\mathbf{n} \\ 0 & 1 & \\ 0 & 0 & 1 \end{pmatrix}, \quad (4.4)$$

where  $\mathbf{n} \hat{=} (l_{0,1} \ l_{0,2})^T$  is the unit normal vector of the mean epipolar line  $\mathbf{l}_0$  and  $n$  is a scalar. Points and lines transform as  $\mathbf{m}' = \mathbf{T}_n \mathbf{m}$  and  $\mathbf{l}' = \mathbf{T}_n^{-T} \mathbf{l}$ , respectively. The following proposition characterizes what happens to the point and line interpretations of the eigenvectors of the epipolar line covariance matrix in the original coordinate frame as the image is translated by  $\mathbf{T}_n$  and  $n \rightarrow \infty$ .

**Proposition 4.3.1** *Let  $\mathbf{l}_0 \neq \mathbf{l}_\infty$  be the estimated epipolar line corresponding to the point  $\mathbf{m}$  in the first view. Translate the second view by the translation matrix  $\mathbf{T}_n$  and let the epipolar lines be parameterized by  $\mathbf{l}'(\mathbf{m}) = \|\mathbf{T}_n^{-T} \mathbf{F} \mathbf{m}\|^{-1} \mathbf{T}_n^{-T} \mathbf{F} \mathbf{m}$  in the translated coordinate system. Let the corresponding epipolar line covariance matrix be  $\mathbf{C}_{Y,n}$  and assume that it has the rank of 2 with no multiple eigenvalues. Denote the ordered unit eigenvectors by  $\mathbf{u}'_{i,n}$ , their point interpretations in the original coordinate frame by  $\mathbf{u}^p_{i,n}$ , and line interpretations by  $\mathbf{u}^l_{i,n}$ ,  $i = 1, 2, 3$ . Assume, however, that the point sequence  $\{\mathbf{u}^p_{2,n}\}$  converges to a finite limit point*



$\mathbf{u}_2^p$  as  $n \rightarrow \infty$ . Then, the corresponding other point and line sequences converge to the following limits in the original coordinate frame.

$$\begin{aligned}
 \mathbf{u}_1^p &= (\mathbf{d}^T \ 0)^T && \text{(the least probable point match on the mean epipolar line)} \\
 \mathbf{u}_2^p &\triangleq \mathbf{T}_u^{-1} \mathbf{e}_3 && \text{(the most probable match in the second image)} \\
 \mathbf{u}_3^p &= (\mathbf{n}^T \ 0)^T && \text{(the least probable pencil of epipolar lines)} \\
 \mathbf{u}_1^l &\triangleq \mathbf{T}_u^T (\mathbf{d}^T \ 0)^T && \text{(the least probable epipolar line that intersects } \mathbf{u}_2^p) \\
 \mathbf{u}_2^l &= \mathbf{l}_\infty && \text{(the least probable epipolar line that intersects } \mathbf{u}_1^p) \\
 \mathbf{u}_3^l &\triangleq \mathbf{T}_u^T (\mathbf{n}^T \ 0)^T \triangleq \mathbf{l}_0 && \text{(the mean epipolar line corresponding to } \mathbf{m}),
 \end{aligned}$$

where  $\mathbf{d} \triangleq (-l_{0,2} \ l_{0,1})^T$  is the unit direction vector of the mean epipolar line and  $\mathbf{T}_u$  is the translation matrix that brings  $\mathbf{u}_2^p$  to the origin, i.e.,  $\mathbf{e}_3 \triangleq \mathbf{T}_u \mathbf{u}_2^p$ .

*Proof.* For convenience, let us define the normalizing transform  $\mathbf{q} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  as

$$\mathbf{q}(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}. \quad (4.5)$$

Let us additionally define  $\mathbf{T}_v$  as an arbitrary finite translation

$$\mathbf{T}_v = \begin{pmatrix} 1 & 0 & \mathbf{v} \\ 0 & 1 & \\ 0 & 0 & 1 \end{pmatrix}, \quad (4.6)$$

where  $\mathbf{v} \in \mathbb{R}^2$ . Since translation transforms  $\mathbf{q}(\mathbf{T}_v \mathbf{u})$  (for points), and  $\mathbf{q}(\mathbf{T}_v^{-T} \mathbf{u})$  (for lines) are continuous and invertible on the unit sphere, the convergence of a point or line sequence in the translated domain is equivalent to the convergence in the original domain.

Let us translate the image by first  $\mathbf{T}_v$  and then by  $\mathbf{T}_n$ . The mean epipolar line transforms as

$$\mathbf{l}'_0 = \mathbf{q}(\mathbf{T}_n^{-T} \mathbf{T}_v^{-T} \mathbf{l}_0). \quad (4.7)$$

As we compute the epipolar line covariance matrix  $\mathbf{C}_{l',n}$ ,  $\mathbf{l}'_0 = \mathbf{u}'_{3,n}$  according to Proposition 4.2.1. Hence,

$$\mathbf{u}_3^p = \lim_{n \rightarrow \infty} \mathbf{q}(\mathbf{T}_v^{-1} \mathbf{T}_n^{-1} \mathbf{u}'_{3,n}) = \lim_{n \rightarrow \infty} \mathbf{q}\left((\mathbf{T}_v^T \mathbf{T}_n^T \mathbf{T}_n \mathbf{T}_v)^{-1} \mathbf{l}_0\right) = \begin{pmatrix} \mathbf{n} \\ 0 \end{pmatrix}, \quad (4.8)$$

as a straightforward calculation shows. Similarly, by assuming  $\{\mathbf{u}_{2,n}^p\}$  converges to a finite point and using the continuity of the mapping,

$$\begin{aligned}
 \mathbf{u}_1^p &= \lim_{n \rightarrow \infty} \mathbf{q}(\mathbf{T}_v^{-1} \mathbf{T}_n^{-1} \mathbf{u}'_{1,n}) \\
 &= \lim_{n \rightarrow \infty} \mathbf{q}(\mathbf{T}_v^{-1} \mathbf{T}_n^{-1} ((\mathbf{T}_n \mathbf{T}_v \mathbf{u}_{2,n}^p) \times (\mathbf{T}_n^{-T} \mathbf{T}_v^{-T} \mathbf{l}_0))) = \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix}.
 \end{aligned} \quad (4.9)$$

For the line interpretation  $\mathbf{u}_2^l$  we get with similar assumptions

$$\begin{aligned} \mathbf{u}_2^l &= \lim_{n \rightarrow \infty} \mathbf{q}(\mathbf{T}_v^T \mathbf{T}_n^T \mathbf{u}_{2,n}') \\ &= \lim_{n \rightarrow \infty} \mathbf{q}(\mathbf{T}_v^T \mathbf{T}_n^T ((\mathbf{T}_n^{-T} \mathbf{T}_v^{-T} \mathbf{l}_0) \times ((\mathbf{T}_n \mathbf{T}_v \mathbf{u}_{2,n}^p) \times (\mathbf{T}_n^{-T} \mathbf{T}_v^{-T} \mathbf{l}_0)))) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \end{aligned} \quad (4.10)$$

By assumption  $\{\mathbf{u}_{2,n}^p\}$  converges to a finite limit, hence  $\exists N \in \mathbb{N}$  such that  $n > N \implies \mathbf{u}_{2,n}^p$  is finite. Since  $\mathbf{T}_v$  is an arbitrary finite translation, for  $n > N$  we may set  $\mathbf{v} = -(u_{32,n}^p)^{-1}(u_{12,n}^p \ u_{22,n}^p)^T$  and denote the corresponding translation matrix by  $\mathbf{T}_{u_n}$  and its limit by  $\mathbf{T}_u$ . Now  $\mathbf{T}_u \mathbf{u}_2^p \hat{=} \mathbf{e}_3 \Leftrightarrow \mathbf{u}_2^p \hat{=} \mathbf{T}_u^{-1} \mathbf{e}_3$  and

$$\begin{aligned} \mathbf{u}_1^l &= \lim_{n \rightarrow \infty} \mathbf{q}(\mathbf{T}_{u_n}^T \mathbf{T}_n^T \mathbf{u}_{1,n}') \\ &= \mathbf{q}\left(\lim_{n \rightarrow \infty} \mathbf{T}_{u_n}^T \lim_{n \rightarrow \infty} \mathbf{q}(\mathbf{T}_n^T ((\mathbf{T}_n \mathbf{T}_{u_n} \mathbf{u}_{2,n}^p) \times (\mathbf{T}_n^{-T} \mathbf{T}_{u_n}^{-T} \mathbf{l}_0))))\right) \\ &= \mathbf{q}\left(\mathbf{T}_u^T \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix}\right) \hat{=} \mathbf{T}_u^T \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix}. \end{aligned} \quad (4.11)$$

Since  $\mathbf{l}_0^T \mathbf{u}_2^p = 0$ , we may also write  $\mathbf{l}_0 \hat{=} \mathbf{T}_u^T (\mathbf{n} \ 0)^T$ .

The proposition thus suggests that translating the origin far away from the image corresponds in the original coordinate frame to the assumption that the least probable epipolar line approaches the line at infinity, and the least probable epipolar line that intersects the most probable point  $\mathbf{u}_2$  is perpendicular to the mean epipolar line, as summarized in Table 4.2. In addition, the least probable point of the mean epipolar line is its ideal point. If we set the origin to  $\mathbf{u}_2^p$  in the original coordinate frame, we would have  $\mathbf{T}_u = \mathbf{I}$  and  $\mathbf{u}_i^p = \mathbf{u}_i^l$ ,  $i = 1, 2, 3$ . Moreover, we would then obtain the same results without translating the origin since the mean epipolar line would be  $\mathbf{u}_3 \equiv \mathbf{l}_0 = (\mathbf{n}^T \ 0)^T$  and  $\mathbf{u}_1 = \mathbf{e}_3 \times (\mathbf{n}^T \ 0)^T = (\mathbf{d}^T \ 0)^T$ . This is intuitively acceptable as well; thus, for practical purposes we recommend the following.

**Recommendation 4.3.2** *When computing the most probable point location or epipolar envelopes, the origin should be translated sufficiently far away from the estimated epipolar line. In this case, the epipolar envelope represents a hyperbola whose asymptotics practically intersect at the most probable matching point.*

By way of example, let us return to a consideration of the case in Fig. 4.3. Let us translate the origin in the normal direction of the estimated epipolar line and observe what happens to the point interpretations in the original coordinate frame. The results are shown in Table 4.3. As one might expect, the most probable point seems to converge to the same limit as the center of the hyperbola

or the epipolar envelope. As the limit seems to be finite,  $\mathbf{u}_1^p$  and  $\mathbf{u}_3^p$  converge to the ideal points according to Proposition 4.3.1. In practice, it seems to suffice that the distance between the origin and the estimated epipolar line is kept at least as a couple of hundred pixels.

**Table 4.2:** Summary of the meaning of the eigenvectors of the epipolar line covariance matrix in the original coordinate frame when the origin is translated far away from the estimated epipolar line  $\mathbf{l}_0$ , corresponding to the point  $\mathbf{m}$  in the first view.

	Interpretation in the image	Illustration
$\mathbf{u}_1^p = \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix}$	The least probable match for $\mathbf{m}$ on the estimated epipolar line $\mathbf{l}_0$ $\Leftrightarrow$ The least probable pencil of epipolar lines that contains $\mathbf{l}_0$	
$\mathbf{u}_1^l \triangleq \mathbf{T}_u^T \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix}$	The least probable epipolar line that intersects the point $\mathbf{u}_2$	
$\mathbf{u}_2^p \triangleq \mathbf{T}_u^{-1} \mathbf{e}_3$	The most probable match for $\mathbf{m}$ in the second image $\Leftrightarrow$ The most probable pencil of epipolar lines in the second image	
$\mathbf{u}_2^l = \mathbf{l}_\infty$	The least probable epipolar line	
$\mathbf{u}_3^p = \begin{pmatrix} \mathbf{n} \\ 0 \end{pmatrix}$	The least probable match for $\mathbf{m}$ in the second image $\Leftrightarrow$ The least probable pencil of epipolar lines	
$\mathbf{u}_3^l \triangleq \mathbf{T}_u^T \begin{pmatrix} \mathbf{n} \\ 0 \end{pmatrix} \triangleq \mathbf{l}_0$	The estimated epipolar line	

**Table 4.3:** The behavior of the point interpretations of the eigenvectors of the epipolar line covariance matrix  $\mathbf{C}_1$  and the center of the epipolar envelope  $\mathbf{c}$  with the 95% confidence level in the original coordinate frame as the origin is translated away in the normal direction of the estimated epipolar line  $\mathbf{l}_0$ . Here  $d$  indicates the amount of translation with respect to the height of the image, which is 512 pixels here.

$d$	$\mathbf{u}_1^p$		$\mathbf{u}_2^p$		$\mathbf{u}_3^p$		$\mathbf{c}$	
	$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$
0	-77.65	178.95	400.93	173.96	0.00	0.00	399.68	173.99
1/4	-233.14	180.57	400.61	173.96	-1.33	-128.00	399.66	173.98
1/2	-471.05	183.05	400.35	173.97	-2.67	-255.99	399.65	173.98
1	-1194.17	190.58	400.03	173.97	-5.33	-511.97	399.65	173.98
2	-3629.22	215.95	399.80	173.97	-10.67	-1023.95	399.65	173.98
4	-12453.23	307.90	399.69	173.97	-21.34	-2047.89	399.65	173.98
8	-45915.74	656.57	399.66	173.97	-42.68	-4095.78	399.65	173.98

## 4.4 Epipolar Uncertainty and Matching

Knowing the exact location of the most probable point for correspondence in the second view does not directly reveal the mutual relevance of multiple candidate correspondences in matching. For example, if there is a point near the most probable point but the point is not exactly on the epipolar line, and another point very far away from the most probable point but closer to the mean epipolar line than the first point, it is not clear which one we should select. In this section, we discuss how the covariance matrix  $\mathbf{C}_1$  of the epipolar line can be used to solve this problem.

Let us assume that  $\mathbf{m}$  is a point in the first of two views whose epipolar geometry has been estimated and let the epipolar line covariance matrix  $\mathbf{C}_1$  corresponding to  $\mathbf{m}$  be defined as in (4.1). Let the corresponding degenerate Gaussian density be  $p(\mathbf{l}|\mathbf{l}_0, \mathbf{C}_1)$  whose two eigenvectors, corresponding to non-zero eigenvalues, span the tangent plane  $\pi$  at the mean epipolar line  $\mathbf{l}_0$ ,  $\|\mathbf{l}_0\| = 1$ , on the unit sphere. Then, we may use at least the following principles to characterize the suitability of a match candidate point  $\mathbf{m}'$  in the second view.

- Maximum value of the probability density along the dual line of the candidate point on the dual plane  $\pi$ , i.e.,

$$p_1(\mathbf{m}'|\mathbf{C}_1) = \sup_{\mathbf{m}'^T \mathbf{l} = 0} p(\mathbf{l}|\mathbf{l}_0, \mathbf{C}_1). \quad (4.12)$$

- Tail probability of the Gaussian density outside the contour which has the same probability as the maximum value, i.e.,

$$p_2(\mathbf{m}'|\mathbf{C}_1) = \int_S p(\mathbf{l}|\mathbf{l}_0, \mathbf{C}_1) dS, \quad (4.13)$$

where  $S = \{\mathbf{l} \in \pi \mid p(\mathbf{l}|\mathbf{l}_0, \mathbf{C}_1) < p_1(\mathbf{m}'|\mathbf{C}_1)\}$ .

- Marginal density over the dual line of the candidate point, i.e.,

$$p_3(\mathbf{m}'|\mathbf{C}_1) = \int_L p(\mathbf{l}|\mathbf{l}_0, \mathbf{C}_1) dL, \quad (4.14)$$

where  $L$  means the line  $\mathbf{m}'$  on the dual plane  $\pi$ .

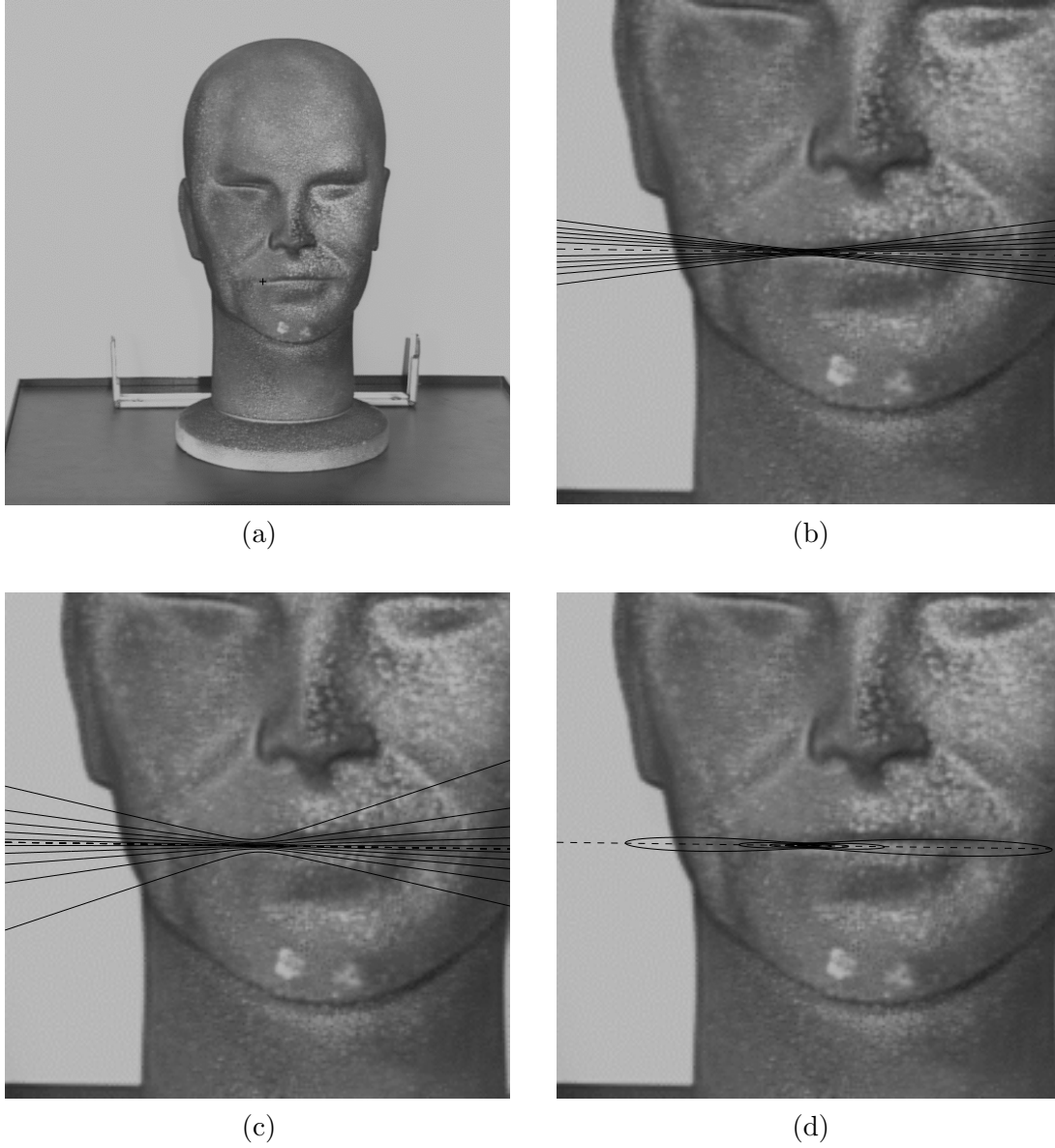
The implementation of these principles should be performed in the transformed dual plane  $\pi'$  for simplicity.

The first principle only means that we take the most probable line from the pencil of lines, defined by  $\mathbf{m}'$ , and use the probability density value of that line as the probabilistic merit of indexing. The second principle is directly related to the error bounds derived by Csurka et al. (1997), because the tail probability gives the confidence level on which the candidate point would intersect the epipolar envelope. The disparity information has not been taken into account because the candidate points may lie on the epipolar lines with an arbitrary distance from the point  $\mathbf{u}_2$ . It is, however, sensible to use either of these two principles, for example when the camera configuration is unchanged but the imaged object is different to those used in epipolar geometry estimation, and thus reveals different disparity distribution.

Following the third principle, we compute the total or marginal probability of all the epipolar lines in the pencil centered at  $\mathbf{m}'$ . In other words, it measures how well the entire pencil of lines supports the estimated epipolar geometry, which is the same criterion that we used in defining the most probable match  $\mathbf{u}_2$ , and therefore the maximum is obtained when  $\mathbf{m}' = \mathbf{u}_2$ . We thus have a probabilistic measure that automatically finds the balance between disparity and orthogonal distance from the epipolar line. This principle is therefore a convenient way to put weights on the candidate matches, particularly when the F-matrix is estimated using the images considered because then, the covariance characterizes the disparities in the particular scene.

Since we can compute the three suitability scores above for any point in the second view, they can be seen as distribution like functions (see Fig. 4.6). In matching, they can be utilized directly by selecting the candidate match on the basis of the best score, or to weight other matching scores before selection, as we will do in the following chapter. However, none of the three principles forms an actual probability density since their integral over the image plane is not unity. Certainly, the integrals of the first two functions diverge, but the third principle may be modified to operate as a proper probability density.

**Proposition 4.4.1** *Let  $\mathbf{m}$  be a point in the first view and assume that the epipolar geometry has been estimated. Let the epipolar line covariance matrix corresponding to  $\mathbf{m}$  be  $\mathbf{C}_1$ , defined as in (4.1) and (4.2), and assume that the sought*



**Figure 4.6:** Contours of the matching probability scores by the three different principles. (a) A point given in the left image. (b) Contours as each pixel in the right image is associated with the probability of the most probable epipolar line which intersect the pixel (4.12). (c) The tail probability in the dual space (4.13), i.e., the contours correspond to  $(1 - k) \times 100\%$  confidence intervals in Csurka *et al.* (1997). (d) Contours of the marginal density over all the lines intersecting a pixel (4.13), i.e., when the disparity information captured by the fundamental matrix covariance is taken into account. The plotted contour values in each image correspond to  $k$  times the corresponding value of the most probable point, where  $k = 0.1, 0.3, 0.5, 0.7, 0.9$ . The estimated epipolar line is shown by the dashed lines. INRIA Syntim owns the copyright of the stereo image pair.

correspondence  $\mathbf{m} \leftrightarrow \mathbf{m}'$  has a priori the same (disparity) distribution as the point matches with which the epipolar geometry has been estimated. Then, up to a Gaussian approximation,  $\mathbf{C}_1$  determines a weak point–point constraint between the views represented by the probability density

$$p(r, \theta | \mathbf{C}_1) = \frac{\sigma_1 \sigma_2 e^{-\frac{1}{2} r^{-2} (\sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta)^{-2}}}{\sqrt{2\pi^3} r^2 (\sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta)^{3/2}}, \quad (4.15)$$

where the point  $\mathbf{m}'$  is parameterized with the signed-distance–direction pair  $(r, \theta)$  on the transformed image plane  $\pi'$ .

*Proof.* In this proof, we use several times the property than when we have an almost everywhere continuous and invertible mapping  $\mathbf{s}: S \rightarrow R$  in any set  $S \subset \mathbb{R}^n$ , we may evaluate the integral of the kernel  $p(\mathbf{s})$  by substitution  $\mathbf{s} = \mathbf{s}(\mathbf{r})$  in the domain  $R$  and it holds

$$\int_S p(\mathbf{s}) d\mathbf{s} = \int_R p(\mathbf{s}(\mathbf{r})) |\det \mathbf{J} \mathbf{s}(\mathbf{r})| d\mathbf{r}, \quad (4.16)$$

where  $\mathbf{J}$  is the Jacobian operator.

Let us transform the original coordinate system by the orthogonal transform  $\mathbf{U}^T$ , where  $\mathbf{U}$  contains the unit eigenvectors of  $\mathbf{C}_1$  in descending order; hence  $\mathbf{m}'' = \mathbf{U}^T \mathbf{m}'$ ,  $\mathbf{e}_3 = \mathbf{U}^T \mathbf{l}_0$ , and  $\mathbf{C}'_1 = \mathbf{U}^T \mathbf{C}_1 \mathbf{U}$  as in (4.2). Let us now consider the dual space of this transformed space. Let us write those lines of the transformed dual space that are parallel to  $\mathbf{m}''$  in parametric form

$$\mathbf{l}_s(t)' = \begin{pmatrix} t\mathbf{d} + s\mathbf{n} \\ 1 \end{pmatrix}, \quad s, t \in \mathbb{R}, \quad (4.17)$$

where  $\mathbf{d}$  is the unit direction vector of the line  $\mathbf{m}''$  and  $\mathbf{n}$  is its unit normal such that the polar angle of  $\mathbf{d}$  is on  $[-\frac{\pi}{2}, \frac{\pi}{2})$  and  $\det(\mathbf{d} \ \mathbf{n}) = 1$ .

By assumption, the epipolar lines, corresponding to the point  $\mathbf{m}$  in the first view, are normally distributed in the transformed dual space with the mean at  $\mathbf{e}_3$  and covariance matrix  $\mathbf{C}'_1$ , and the corresponding density function is  $p(\mathbf{l}_s(t)' | \mathbf{e}_3, \mathbf{C}'_1)$ . The marginal probability of  $p(\mathbf{l}_s(t)' | \mathbf{e}_3, \mathbf{C}'_1)$  over the lines parallel to  $\mathbf{m}''$  depends only on the direction angle  $-\theta^c \in [-\frac{\pi}{2}, \frac{\pi}{2})$  of the line  $\mathbf{m}''$  where  $\theta^c = \frac{\pi}{2} \text{sign } \theta - \theta$  is the complement angle of the direction  $\theta$  of the point  $\mathbf{m}''$ , as we define  $\text{sign } 0 = 1$ . The marginal probability is here denoted by

$$p(s | \theta, \mathbf{C}'_1) = \int_{-\infty}^{\infty} p(\mathbf{l}_s(t)' | \mathbf{e}_3, \mathbf{C}'_1) dt \equiv \int_L p(\mathbf{l} | \mathbf{l}_0, \mathbf{C}_1) dL. \quad (4.18)$$

Let us evaluate the kernel  $p(\mathbf{l}_s(t)' | \mathbf{e}_3, \mathbf{C}'_1)$  or the 2D Gaussian  $p(\tilde{\mathbf{l}}_s(t)' | \mathbf{0}_2, \tilde{\mathbf{C}}'_1)$ , where  $\tilde{\mathbf{l}}_s(t)' = (t\mathbf{d} + s\mathbf{n})$ ,  $\mathbf{0}_2 = (0 \ 0)^T$ , and  $\tilde{\mathbf{C}}'_1 = \text{diag}(\sigma_1^2, \sigma_2^2)$ . Let us make the substitution

$$\begin{pmatrix} t \\ s \end{pmatrix} = \mathbf{R}^T \tilde{\mathbf{C}}_1^{\frac{1}{2}} \mathbf{R}' \begin{pmatrix} u \\ v \end{pmatrix} \Leftrightarrow \tilde{\mathbf{l}}'_s(t) = \tilde{\mathbf{C}}_1^{\frac{1}{2}} \mathbf{R}' \begin{pmatrix} u \\ v \end{pmatrix}, \quad (4.19)$$

where  $\mathbf{R} = (\mathbf{d} \ \mathbf{n})$  and  $\mathbf{R}'^T = (\mathbf{d}' \ \mathbf{n}')^T$  is the rotation that brings the image of the line  $\mathbf{m}''$  horizontal in the  $uv$ -plane, and the rotation angle is on the interval  $[-\frac{\pi}{2}, \frac{\pi}{2})$ . Now,

$$\begin{aligned} p(\tilde{\mathbf{l}}_s(t)' | \mathbf{0}_2, \tilde{\mathbf{C}}'_1) &\hat{=} |\det(\mathbf{R}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}} \mathbf{R}')| \ p\left(\tilde{\mathbf{C}}_1'^{\frac{1}{2}} \mathbf{R}' \begin{pmatrix} u \\ v \end{pmatrix} \middle| \mathbf{0}_2, \tilde{\mathbf{C}}'_1\right) \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(u^2+v^2)}, \end{aligned} \quad (4.20)$$

where  $\hat{=}$  denotes the equivalence between the two kernels. Hence,

$$\int_{-\infty}^{\infty} p(\mathbf{l}_s(t)' | \mathbf{e}_3, \mathbf{C}'_1) dt \hat{=} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+v^2)} du = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}. \quad (4.21)$$

Since  $0 = \mathbf{n}^T \mathbf{d} = \mathbf{n}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}} \tilde{\mathbf{C}}_1'^{-\frac{1}{2}} \mathbf{d} = \frac{\mathbf{n}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}}}{\|\mathbf{n}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}}\|} \frac{\tilde{\mathbf{C}}_1'^{-\frac{1}{2}} \mathbf{d}}{\|\tilde{\mathbf{C}}_1'^{-\frac{1}{2}} \mathbf{d}\|} = \mathbf{n}'^T \mathbf{d}'$ ,  $\tilde{\mathbf{C}}_1'^{-\frac{1}{2}}$  preserves the orientation, and  $\mathbf{d}'$  lies in the same quadrant as  $\mathbf{d}$ , we return to the original coordinate frame by the substitution

$$v = \mathbf{n}'^T \tilde{\mathbf{C}}_1'^{-\frac{1}{2}} \tilde{\mathbf{l}}'_s(t) = \frac{\mathbf{n}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}}}{\|\mathbf{n}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}}\|} \tilde{\mathbf{C}}_1'^{-\frac{1}{2}} \tilde{\mathbf{l}}'_s(t) = \frac{s}{\|\mathbf{n}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}}\|}. \quad (4.22)$$

By denoting that  $\mathbf{n} = (-\sin -\theta^c \ \cos -\theta^c)^T = \text{sign } \theta (\cos \theta \ \sin \theta)^T$  and  $\|\mathbf{n}^T \tilde{\mathbf{C}}_1'^{\frac{1}{2}}\|^2 = \sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta$ , we get

$$p(s|\theta, \mathbf{C}'_1) = \left| \frac{dv}{ds} \right| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} = \frac{e^{-\frac{1}{2}s^2(\sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta)^{-1/2}}}{\sqrt{2\pi(\sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta)}} \quad (4.23)$$

that is a 1D Gaussian.

The lines parallel to  $\mathbf{m}''$  intersect at  $(\mathbf{d} \ 0)^T$  in the transformed dual space that is equivalently the line that joins the origin  $\mathbf{e}_3$  and  $\mathbf{m}''$  in the transformed image space. In addition, we have parameterized the parallel lines in the transformed dual space by  $s$  which represents the signed distance between the line  $\mathbf{m}''_s$  and  $\mathbf{e}_3$  in the transformed dual space. As we define  $r$  as the signed distance from  $\mathbf{e}_3$  in the transformed space, it also parameterizes the points of line  $(\mathbf{d} \ 0)^T$  uniquely. Since the lines  $\mathbf{m}''_s$  and points on the line  $(\mathbf{d} \ 0)^T$  have one-to-one correspondence, we have  $s = s(r)$ . In fact,

$$s = -\frac{\text{sign}\left(\frac{m''_3}{m''_2}\right) |\mathbf{m}''^T \mathbf{e}_3|}{\sqrt{m''_1{}^2 + m''_2{}^2}} = -\frac{\text{sign}\left(\frac{m''_3}{m''_2}\right)}{\sqrt{\left(\frac{m''_1}{m''_3}\right)^2 + \left(\frac{m''_2}{m''_3}\right)^2}} \quad (4.24)$$

and

$$r = \text{sign}\left(\frac{m''_1}{m''_3}\right) \sqrt{\left(\frac{m''_1}{m''_3}\right)^2 + \left(\frac{m''_2}{m''_3}\right)^2}, \quad (4.25)$$



hence,

$$s(r) = \begin{cases} 1/r, & \theta \in [-\frac{\pi}{2}, 0) \\ -1/r, & \theta \in [0, \frac{\pi}{2}) \end{cases}. \quad (4.26)$$

Therefore the conditional probability density

$$p(r|\theta, \mathbf{C}_1) = \frac{1}{r^2} p(s(r)|\theta, \mathbf{C}'_1) = \frac{e^{-\frac{1}{2}r^{-2}(\sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta)^{-1/2}}}{r^2 \sqrt{2\pi(\sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta)}}. \quad (4.27)$$

Let us then consider the probability density  $p(\theta|\mathbf{C}_1)$  of the direction angle  $\Theta = -(-\Theta^c)^c = -\operatorname{arccot} L_2/L_1 = -\arctan L_1/L_2$ , where the random vector  $\mathbf{L} = (L_1 \ L_2 \ L_3)^T \sim N(\mathbf{e}_3, \mathbf{C}'_1)$ . Let us make the substitution

$$\mathbf{l} \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \begin{pmatrix} \sigma_1 r' \cos \theta' \\ \sigma_2 r' \sin \theta' \end{pmatrix}, \quad (4.28)$$

where  $r' \in \mathbb{R}$ ,  $\theta' \in [-\frac{\pi}{2}, \frac{\pi}{2})$ . The determinant of the Jacobian of this mapping is  $\sigma_1 \sigma_2 r'$ : hence, as above we get

$$p(r', \theta'|\mathbf{C}'_1) = \sigma_1 \sigma_2 |r'| p(\mathbf{l}(r', \theta')|\mathbf{e}_3, \mathbf{C}'_1) = \frac{|r'|}{2\pi} e^{-\frac{1}{2}r'^2}. \quad (4.29)$$

The marginal probability over  $r'$  is

$$p(\theta'|\mathbf{C}'_1) = \int_{-\infty}^{\infty} p(r', \theta'|\mathbf{C}'_1) dr' = \frac{1}{\pi} \int_0^{\infty} r' e^{-\frac{1}{2}r'^2} dr' = \frac{1}{\pi}. \quad (4.30)$$

We may now transform  $\theta'$  back by the substitution  $\theta' = -\arctan(\frac{\sigma_1}{\sigma_2} \cot \theta)$ , when the density function for  $\theta$  is obtained as

$$p(\theta|\mathbf{C}_1) = p(\theta'|\mathbf{C}'_1) \left| \frac{d\theta'}{d\theta} \right| = \frac{\sigma_1 \sigma_2}{\pi(\sigma_1^2 \cos^2 \theta + \sigma_2^2 \sin^2 \theta)}. \quad (4.31)$$

The joint probability density  $p(r, \theta|\mathbf{C}_1)$  is now obtained by combining the results above, thus,

$$p(r, \theta|\mathbf{C}_1) = p(r|\theta, \mathbf{C}_1) p(\theta|\mathbf{C}_1), \quad (4.32)$$

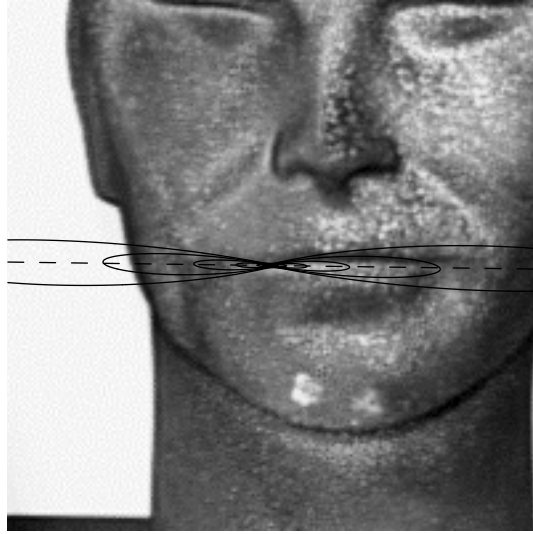
and the claim follows.

**Corollary 4.4.2** *The probability density in (4.15), represented by  $\mathbf{m}' = (x' \ y' \ 1)^T$  in coordinate basis of the second image, is*

$$p(x', y'|\mathbf{C}_1) = |\det \mathbf{J} \ \mathbf{r}(x', y')| \ p(\mathbf{r}(x', y')|\mathbf{C}_1), \quad (4.33)$$

where

$$\mathbf{r} \begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} \operatorname{sign} \left( \frac{\mathbf{u}_1^T \mathbf{m}'}{\mathbf{u}_3^T \mathbf{m}'} \right) \sqrt{\left( \frac{\mathbf{u}_1^T \mathbf{m}'}{\mathbf{u}_3^T \mathbf{m}'} \right)^2 + \left( \frac{\mathbf{u}_2^T \mathbf{m}'}{\mathbf{u}_3^T \mathbf{m}'} \right)^2} \\ \arctan \frac{\mathbf{u}_2^T \mathbf{m}'}{\mathbf{u}_1^T \mathbf{m}'} \end{pmatrix}. \quad (4.34)$$



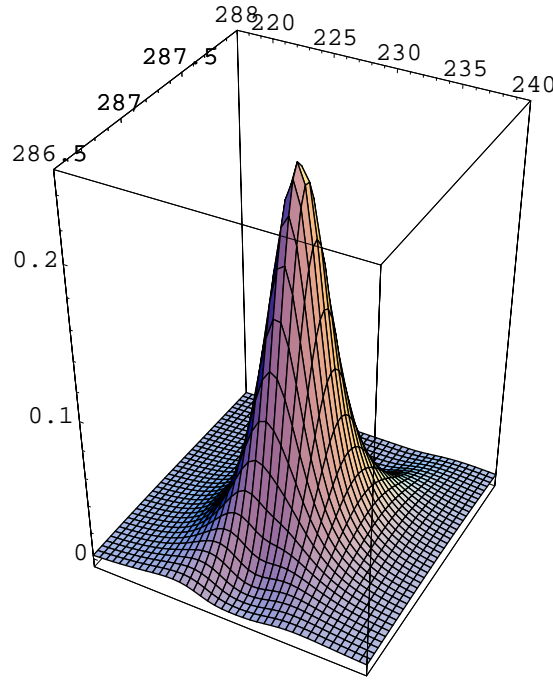
**Figure 4.7:** Equi-probability contours of the probability density (4.33) in the second image, given the point shown in Fig. 4.6a, at the levels of  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$ , and  $10^{-1}$  times the maximum value at  $\mathbf{u}_2$ . The probability is strongly peaked around  $\mathbf{u}_2$  (cf. Fig. 4.8).

For comparison with Fig. 4.6, we plotted contours of the derived probability density in Fig. 4.7, and the density is displayed in Fig. 4.8. As the contours are propeller-shaped where the narrowest location is around the point  $\mathbf{u}_2$ , it is improbable that a match would reside near  $\mathbf{u}_2$  but deviated in the normal direction of the mean epipolar line. On the other hand, the contours additionally determine the plausible deviation in the direction of the epipolar lines. Here, the density is peaked around  $\mathbf{u}_2$  strongly since the depth variation is relatively small in the scene.

We have thus derived a *weak* point–point constraint between two views or, in other words, a closed-form conditional distribution for a match in the other view given the training samples used in the epipolar geometry and its uncertainty estimation. We assumed that the epipolar lines are normally distributed in the dual space, but similar derivation could be performed with certain other density assumptions. The derived density can be directly utilized in matching and it is a most convenient tool for computing, for instance, the dense disparity map of the scene with Bayes methods.

## 4.5 Summary and Discussion

In this chapter, we have investigated the properties of the epipolar line covariance matrix and discussed its practical utilization. We first showed that the eigenvec-



**Figure 4.8:** Probability density (4.33) in the second image, given the point shown in Fig. 4.6a. The probability is strongly peaked around  $\mathbf{u}_2$  because the depth variation is relatively small in the scene. Note the scaling of the axes.

tors of the covariance matrix have coupled point and line interpretations in the other image. In fact, given a point in one image, the most probable matching location in the other image is represented by the second eigenvector, in the sense that represents the pencil of epipolar lines with the largest total probability. Correspondingly, the least probable point on the estimated epipolar line is the first eigenvector, and the third eigenvector represents the least probable point of all points.

The eigenvectors similarly have line interpretations and, in essence, the third eigenvector is the estimated epipolar line, which implies that the location of the image origin affects their realizations as well as the shape of epipolar envelopes. This is clearly an undesired property, so we proposed that the origin should be translated far away from the estimated epipolar line. In this way, we obtain intuitively acceptable interpretations: the least probable point on the estimated epipolar line is its ideal point, the least probable point of all points is also an ideal point but in the normal direction of the estimated epipolar line, and the least probable epipolar line is the line at infinity. While we did not prove the assumed convergence of the point interpretation sequence of the second eigenvector in the

original frame, the convergence seems to take place in practice.

The derivation of the weak point–point probability density from the epipolar line covariance matrix, which we consider an important result, completes this chapter. Although the result does not represent a strong geometric constraint between the views, it is a closed-form representation of the conditional probability density for the correspondence, given a point in the first view and the training data that has been used in the epipolar geometry estimation. Since the probability density has been derived from the Gaussian approximation in the dual space, it has the advantages of the corresponding Triggs' (2001) joint feature distributions. However, we do not have to constrain ourselves to linear estimation but may use the appropriate statistical criteria and robust methods in estimating the epipolar geometry.

## Chapter 5

# Multi-Resolution Image Matching<sup>1</sup>

The most important applications of epipolar geometry are image matching and feature point tracking since the known epipolar geometry reduces the matching dimensionality from two to one. In the application part of this thesis (Chapter 8), we need a reliable method for tracking points in electron microscope images; hence, we propose a wavelet-based technique for matching points in uncalibrated images, where the matching principle is simple and efficient. The major contribution of this chapter is showing that an appropriate use of multi-resolution information provides fundamental cues for discovering false matches. In addition, if multiple match candidates occur, we select the most probable using the covariance information of the fundamental matrix, as proposed in the previous chapter. As far as image matching is concerned, the experimental results show that the false match probability practically vanishes when natural images are used.

### 5.1 Introduction

Stereo matching and feature point tracking are important topics in computer vision and are accordingly widely researched. A common problem in establishing correspondence between two images is that the estimated point-to-point matches consist of many false matches. This is because the interest points are normally matched by direct correlation of their neighborhoods (Zhang et al. 1994, Hartley and Zisserman 2000), but the correlation window often matches well to many incorrect locations in the other image. To alleviate the matching problem, numerous multi-resolution approaches have been proposed in recent decades, (see e.g. Marr and Poggio 1979, Hannah 1989, Hoff and Ahuja 1989, Weng, Huang

---

<sup>1</sup>A short version of this chapter has been published in Brandt and Heikkonen (2001a).

and Ahuja 1992, Wei, Brauer and Hirzinger 1998, Liu and Bhattacharya 2000, Alvarez, Deriche, Sánchez and Weickert 2000).

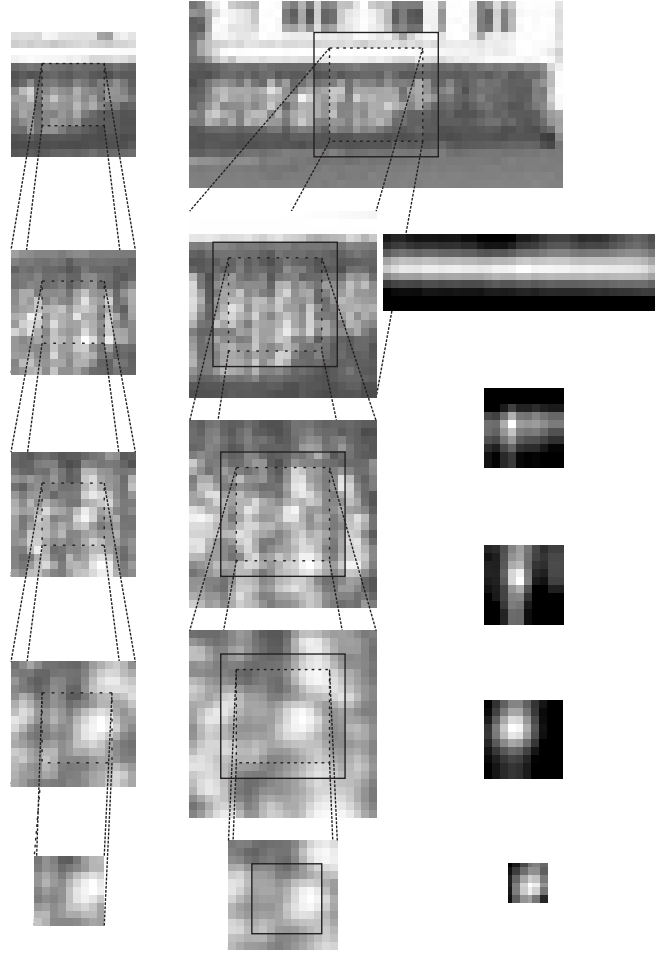
In this chapter, we construct a more reliable multi-resolution based interest point matching instead of using only a one-sized correlation window. In contrast to typical previous multi-resolution matching work, our goal (as well as in Lew and Huang 1999) is not to compute a dense disparity map for images but rather to match reliably and efficiently two sets of discrete interest points, here detected by the improved Harris corner detector (Schmid, Mohr and Bauckhage 2000, Harris and Stephens 1988). The weak point-point probability distribution, derived in the previous chapter, can also be easily utilized; hence, the balance between the disparity and the deviation from the epipolar line can be automatically determined. As our experiments show, we finally have an efficient and reliable matching algorithm that is practically free from mismatches.

## 5.2 Wavelet-Based Matching

Suppose now that the fundamental matrix  $\mathbf{F}$  and its covariance matrix  $\mathbf{C}_\mathbf{F}$  have been estimated. The problem remains to find the matching feature points, here corners, using the epipolar geometry. The matching method we propose is motivated by the way humans seem to perform the task. In short, one would first look for similarity at a lower resolution level, then zoom to the details at a higher resolution, and go on to the highest resolution level if the similarity is good enough. If the match is not good at higher resolution, humans would go back to the lower resolution and search for a new candidate. This idea is implemented as follows.

Firstly, the images are rectified (Faugeras 1993, Hartley and Zisserman 2000), i.e., the images are transformed so that the epipolar lines are horizontal and vertically aligned with respect to each other. After the rectification, wavelet decomposition (for a great introduction, see Burrus, Gopinath and Guo 1998) is computed for both images with the Haar basis. The wavelet representation can be thought of as an image pyramid, in which the original image has multiple resolution levels, with the highest resolution image on the bottom and the coarsest on the top. We decomposed the images to four resolution levels (original + three lower resolution levels). If the images consisted of several repeating patterns, the decomposition could naturally be performed even more times, but in our case this was enough.

For every corner point in the left rectified image we do the following. As a search window in the right image, we select a rectangular area on the lowest resolution image, centered at the most probable point (see Section 4.2) on the epipolar line. The correspondence is to be sought on this area. In practice, we thus truncate the searching area and assume bounded disparity. For the size of the search window, we set  $48 \times 24$ , which corresponds to a window of  $384 \times 192$  pixels in the



**Figure 5.1:** Multi-resolution template matching with Haar wavelets. Left column: neighborhood templates at four scales where the lower resolution templates represent larger neighborhood. Middle column: the search regions in the second image. Right column: correlation images.

original image. Around the considered corner point in the left image, we extract a template window of  $16 \times 16$  at each resolution level (see Fig. 5.1).

Starting from the lowest resolution Level, we compute the standard correlation coefficient  $\rho$  (Gonzalez and Woods 1993) between the search window and the template at each possible location. The result is the correlation image of size  $33 \times 9$ . Then, we search the local maxima on the 8-neighborhood in the correlation image and save all those locations to a tree data structure with the correlation score where the correlation is statistically significant ( $\rho > 0.8$ ).

Next, we consider one level higher resolution and cut  $24 \times 24$  window from each found maximum correlation location. Inside this window, we again compute the

correlation coefficient and search for the local maxima. The found locations and correlation scores are correspondingly saved in the tree under the node considered. This procedure is recursively performed for all the significant correlation maxima. Finally, if we reach the highest resolution level, we perform the computation with a  $9 \times 9$  window in the left image and a  $14 \times 14$  window in the right image in order to refine corner point location.

When the above is done, we remove from the tree all the candidates which do not lie inside the 95% confidence intervals of the corresponding epipolar lines (Csurka et al. 1997). We further remove all the candidates whose Euclidean distance from the nearest corner point (in the right image) is over  $\sqrt{2}$ . Performing in this way, in most cases we obtain at most one candidate whose coordinates are identified with the nearest corner point. Moreover, since we use the information on multiple scales, the probability of obtaining false matches is practically negligible with most natural images. If, however, multiple candidates occur, we propose selecting the one which maximizes the average correlation score  $\rho_w$  over all the resolution levels, weighted by the criterion<sup>2</sup> (4.33), i.e.,

$$\rho_w = p(\mathbf{m}' | \mathbf{C}_1) \bar{\rho}. \quad (5.1)$$

## 5.3 Experiments

To illustrate how the proposed method works in practice, we matched four stereo image pairs whose left images are shown in Fig. 5.2. In these experiments, we used the affine approximation for the fundamental matrix, which was estimated with the covariance matrix, as proposed in Chapter 3. From the aerial stereo pair, we found 1640 corners from the left image and 1986 from the right, of which 215 were found to correspond. As Figure 5.3a indicates, no mismatches were found.

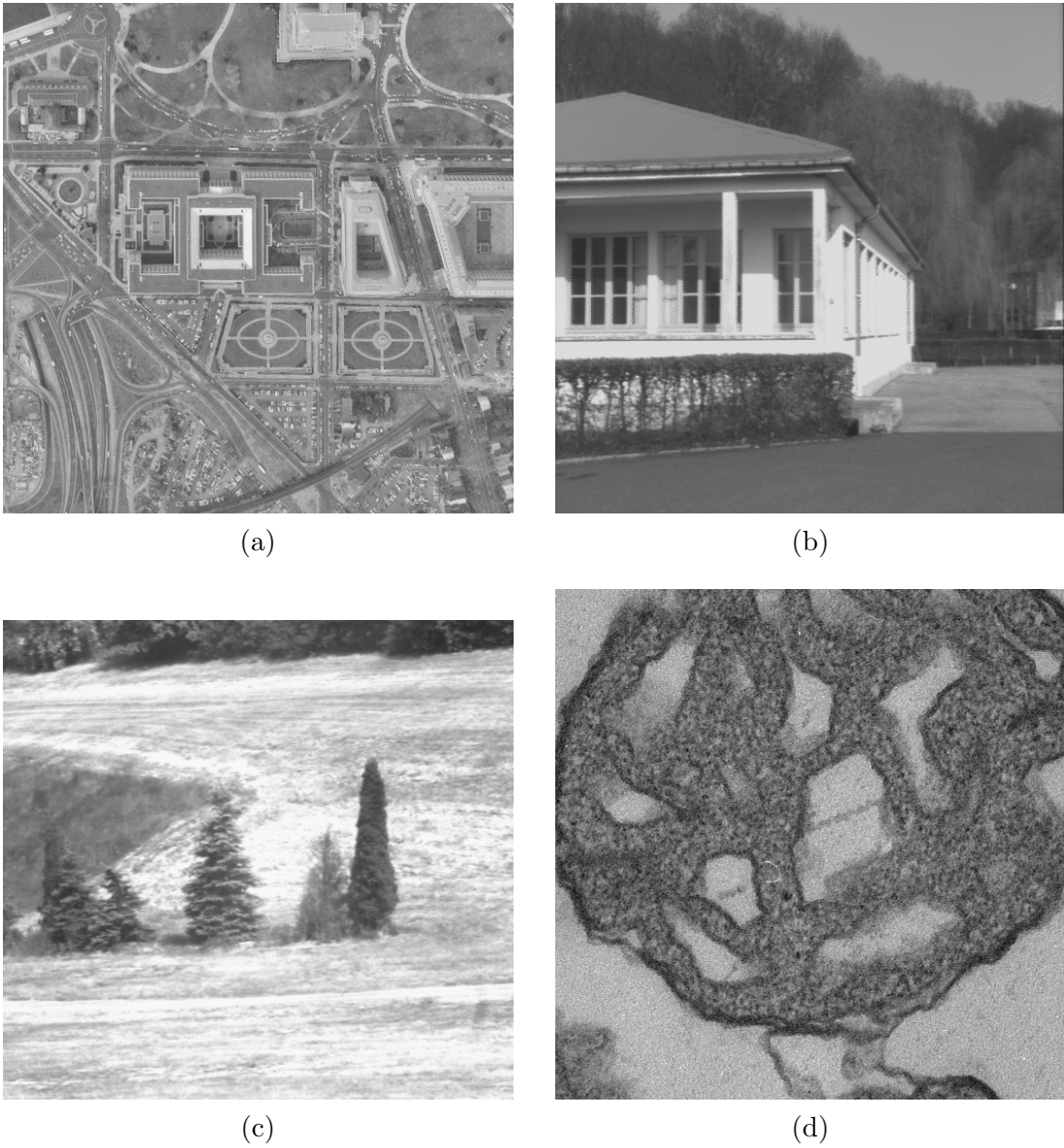
From the left image of the house pair, 583 corner points were found 903 from the right. The matching algorithm succeeded to associate 178 corner pairs. From these pairs only one mismatch was identified, and it is shown in Fig. 5.3b. The mismatch results from matching some suspicious corner points on the edge of the pillar. However, as can be seen from Fig. 5.3b, even difficult small details such as the branches in the bushes have been matched correctly.

From the stereo image pair of trees, 621 corners were found in the left image and 434 from the right. The matching resulted 287 correspondences and their flow is illustrated in Fig. 5.3c. As can be seen, there are no mismatches. As the fourth image pair, we had noisy images representing a mitochondrion section. Because of the large size of the images, the Harris corner detector resulted 7417 corner

---

<sup>2</sup>Because the criterion (4.33) is the most recent result of this thesis, we used the criterion (4.14) in the experiments. Nevertheless, this is hardly significant since multiple candidates were very rare in practice and the two criteria are somewhat similar.





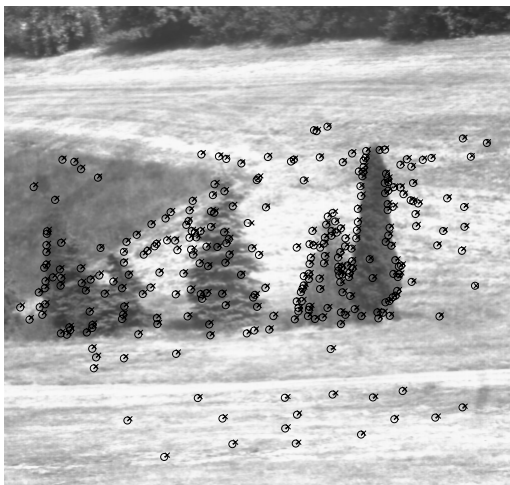
**Figure 5.2:** Left images of the stereo pairs used in experiments. (a) Aerial view of a town; (b) house; (c) trees; (d) slice of a mitochondrion. The aerial view and the tree image pairs are from the CMU VASC Image Database; copyrights of the house pair belong to INRIA-Syntim.



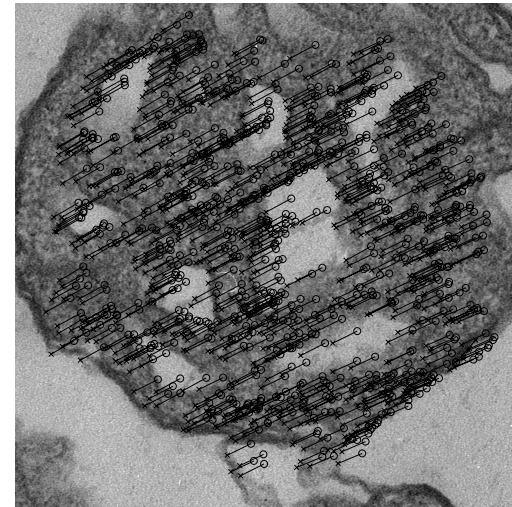
(a)



(b)



(c)



(d)

**Figure 5.3:** Flow of the found correspondences plotted in the right image of the stereo pair. (a) Aerial view (no mismatches); (b) house (one mismatch, pointed by the arrow); (c) trees (no mismatches); (d) slice of the mitochondrion (no mismatches).

points in the left image and 10169 corners in the right image. The matching procedure thereafter found correspondence between 801 corner pairs. Also, in this case, no clear mismatches were found (see Fig. 5.3d).

## 5.4 Summary

In this chapter, we have proposed a multi-resolution method for matching points in two images. By combining the use of epipolar geometry, its uncertainty and multi-resolution intensity information, a robust and efficient method can be obtained. The experiments indicate that the method is reliable as the probability for obtaining false matches practically vanishes, especially when natural images are used. The method should therefore be well suitable for point tracking applications.



## Chapter 6

# Affine Reconstruction and Missing Data<sup>1</sup>

As we have established point matches over several views, the next problem is to reconstruct the points in the 3-space and solve the camera projection matrices simultaneously up to an overall affinity. Assuming an affine camera model, the problem can be solved by the factorization algorithm that gives the maximum likelihood solution for affine reconstruction. It requires, however, that all the feature points be visible in all views. Therefore, we derive here a closed-form-expression for the 3D coordinates of the feature points and translation vectors given the inhomogeneous affine projection matrices, but no single feature point needs to be visible in all views. The expression represents closed-form maximum likelihood affine triangulation under missing data and unknown translations, and it implies two iterative algorithms for the maximum likelihood affine reconstruction, where all the measured data may be used. The solution additionally has applications in affine bundle adjustment, used in Chapters 7 and 8, and identifying degenerate configurations.

### 6.1 Introduction

A central problem in computer vision is the structure-from-motion problem, whereby both the structure of the scene and the motion of the camera are to be solved from image measurements. Despite the considerable attention that has been given to the problem, it has not yet been completely solved. The problem can be approached from different types of correspondences and assumptions for the camera model. As the correspondence between the images must be established, a typical way is to use point correspondences as considered here. Other

---

<sup>1</sup>The content of this chapter has been partly published in Brandt (2002a).

possibilities are line (Weng et al. 1992, Quan and Kanade 1996) or conic (Ma 1993, Kahl and Heyden 1998) correspondences.

If the full perspective projection camera model were selected, the reconstruction problem with missing data would be particularly difficult. In fact, all the known methods that minimize the geometrical error in the projective reconstruction are iterative (Hartley and Zisserman 2000) even though missing data were not considered. The full projective reconstruction problem corresponds to the bundle adjustment problem that is widely researched (for a review, see Triggs, McLauchlan, Hartley and Fitzgibbon 2000). Because of its difficulty and nonlinearity, the image set is frequently split into parts that are thereafter solved separately, and the results are combined in a more or less heuristic way.

In contrast to the full perspective projection model, the non-iterative solution for the affine reconstruction may be obtained using singular value decomposition. In fact, the well-known factorization algorithm proposed by Tomasi and Kanade (1992) and Poelman and Kanade (1997) achieves the maximum likelihood affine reconstruction under the assumption of isotropic, mean-zero Gaussian noise, as noted by Reid and Murray (1996). It is thus, in this sense, an optimal method if all the feature points are visible in each view. Even some assumptions of noise covariance can be taken into account, as proposed by Irani and Anandan (2000); an iterative algorithm incorporating uncertainty models is given by Morris and Kanade (1998). Robust factorization, based on LMedS, has also been proposed (Huynh and Heyden 2001). However, problems arise if there are missing measurements, as is almost always the case in realistic feature point tracking with several images.

So far, many heuristics have been attempted when dealing with missing data. Tomasi and Kanade (1992) first proposed that the missing data could be sequentially replaced using complete subsets of the data. In addition however, as pointed out by Jacobs (2001), the problem of finding the largest submatrix with missing elements is NP-hard, it is not clear how the inaccuracies will propagate in the computation of additional missing elements. Jacobs also demonstrated cases that cannot be solved by Tomasi and Kanade's imputation algorithm. Jacobs (2001) in turn proposes a non-iterative method, whereby the measurement matrix is constrained to satisfy certain rank constraints under absent measurement values. The result of the Jacobs' method is suboptimal but reported to be useful as an initial estimate for iterative methods. A corresponding approach is proposed by Kahl and Heyden (1999). Shum, Ikeuchi and Reddy (1995) propose a weighted least squares method that is a correct formulation for iteratively searching the ML solution but works only without translations or if the translations are known. The solution is bilinear, following the formulation of Wiberg (1976). A similar approach is Morris and Kanade's (1998) if one sets infinite variance for missing points and unity for others; in this case, however, their proposed method for computing the translations breaks down, as we will observe in Section 6.3.

If we assume the affine camera model and allow that each feature point is not shown in every image, the optimal values for the translation parameters may be easily derived as will be seen in Section 6.3; the main result of this chapter is, however, obtained by carrying out a corresponding analysis on the three-dimensional feature point coordinates to be reconstructed. As presented in Section 6.4, we have derived the closed-form maximum likelihood solution for all the feature point coordinates in the world coordinate frame, including translations, given the inhomogeneous affine projection matrices. In other words, this result implies maximum likelihood closed-form affine triangulation under missing data. The result is additionally globally optimal since all views and feature points are used simultaneously.

The solutions were originally derived to help us to solve the registration problem of electron microscope images (see Section 6.5.4), or more generally, to get reasonable initial estimates for the 3D coordinates of the tracked features in the transformation parameter optimization (see Chapters 7 and 8) or bundle adjustment. In addition, the solution can be used to identify degenerate image configurations, as proposed in Section 6.5.3. Another application of the solution is the affine reconstruction with missing data which leads to iterative algorithms for finding the maximum likelihood solution (Section 6.5.2) while all the measurements may be used.

## 6.2 Problem Definition

When rigid motion of a single object or scene is observed in  $N$  images, we may use the following model for the feature point  $j$  in the image  $i$ .

$$\mathbf{m}_j^i = \mathbf{M}^i \mathbf{x}_j + \mathbf{t}^i, \quad (6.1)$$

where  $\mathbf{m}_j^i$  is an inhomogeneous  $2 \times 1$  vector describing the feature point  $j$  coordinates in the image  $i$ ,  $\mathbf{M}^i$  is a  $2 \times 3$  inhomogeneous projection matrix, i.e., a transformation matrix including the projection, scaling, and rotation,  $\mathbf{x}_j$  is the reconstructed 3D coordinate vector of the observed feature point  $j$ , and  $\mathbf{t}^i$  is the translation vector for the image  $i$ .

The goal is to find the least-squares solution for the parameters by minimizing the cost function

$$C(\mathbf{m}) = \sum_i \sum_j (\hat{\mathbf{m}}_j^i - \mathbf{m}_j^i)^T (\hat{\mathbf{m}}_j^i - \mathbf{m}_j^i) \delta_{ij}, \quad (6.2)$$

where  $\mathbf{m}_j^i$  is the measured location for the feature point  $j$  in the image  $i$ ,  $\hat{\mathbf{m}}_j^i$  is its estimated, noise-free counterpart, and  $\delta_{ij}$  is the Kronecker delta product indicating whether the  $j^{\text{th}}$  feature point is found in image  $i$ . In other words, normally distributed noise is assumed at the observed feature point coordinates.

### 6.3 Closed-Form Solution for the Translation Parameters

The least squares solution for the translation parameter can be easily derived with the following result.

**Proposition 6.3.1** *The least squares estimate for the translation parameter, given the other parameters,  $\mathbf{t}^i$  for the  $i^{\text{th}}$  view is obtained from*

$$\hat{\mathbf{t}}^i = \bar{\mathbf{m}}^i - \mathbf{M}^i \bar{\mathbf{x}}^i, \quad (6.3)$$

where

$$\bar{\mathbf{m}}^i = \frac{\sum_j \mathbf{m}_j^i \delta_{ij}}{\sum_j \delta_{ij}} \text{ and } \bar{\mathbf{x}}^i = \frac{\sum_j \mathbf{x}_j \delta_{ij}}{\sum_j \delta_{ij}}.$$

*Proof.* By differentiating (6.2) and setting the result zero it follows:

$$\begin{aligned} \frac{\partial C(\mathbf{m})}{\partial \mathbf{t}^i} &= \mathbf{0} \\ \Leftrightarrow 2 \sum_j (\hat{\mathbf{m}}_j^i - \mathbf{m}_j^i)^T \frac{\partial \hat{\mathbf{m}}_j^i}{\partial \mathbf{t}^i} \delta_{ij} &= \mathbf{0} \\ \Leftrightarrow 2 \sum_j (\mathbf{M}^i \mathbf{x}_j + \mathbf{t}^i - \mathbf{m}_j^i)^T \delta_{ij} &= \mathbf{0} \\ \Leftrightarrow \sum_j \mathbf{t}^i \delta_{ij} &= \sum_j \mathbf{m}_j^i \delta_{ij} - \sum_j \mathbf{M}^i \mathbf{x}_j \delta_{ij} \\ \Leftrightarrow \mathbf{t}^i &= \underbrace{\frac{\sum_j \mathbf{m}_j^i \delta_{ij}}{\sum_j \delta_{ij}}}_{\triangleq \bar{\mathbf{m}}^i} - \mathbf{M}^i \underbrace{\frac{\sum_j \mathbf{x}_j \delta_{ij}}{\sum_j \delta_{ij}}}_{\triangleq \bar{\mathbf{x}}^i} \end{aligned} \quad (6.4)$$

Note that  $M^i \cdot \sum_j \delta_{ij}$  refers to the number of found feature points in image  $i$ . Therefore,  $\bar{\mathbf{m}}^i$  is the center of mass of the found feature points in image  $i$ , and  $\bar{\mathbf{x}}^i$  is the centroid of the corresponding 3D coordinates.

A common result is that the center of gravity of the 3D points projects to the centroid of the points under affine projection. Therefore, if we do *not* have missing data, by setting the 3D origin to the center of mass of all the observed points ( $\sum_j \mathbf{x}_j = \mathbf{0}$ ), the translation parameter for each camera is directly determined. This can similarly be seen from (6.4) by setting  $\delta_{ij} = 1$  that implies  $\mathbf{t}^i = \frac{1}{M} \sum_j \mathbf{m}_j^i$  since the latter term in (6.4) vanishes.

If we have missing data, we may still freely set the origin such that  $\sum_j \mathbf{x}_j = \mathbf{0}$  but the second term in (6.4) is no longer zero in general. In addition, this term



can be seen as a correction term to the centroid of the visible points in image  $i$  to correspond to the projection of the center of mass of all the 3D points. The fundamental issue is that the correction term also depends on the 3D coordinates of the points visible in some image  $i$ , not only their image coordinates, no matter how we fix the 3D origin. Therefore, the actual optimal translation parameter values cannot be determined *prior to* the reconstruction, in contrast to the case when we do not have missing data.

This is the very reason why the previous methods are suboptimal when dealing with missing data. One might think that, for example, factorization with uncertainty models (Morris and Kanade 1998) could solve the missing data problem simply by regarding missing data as any points with infinite variance. In factorization, however, one must deal with object centered coordinates and, with points of infinite variance, their proposed method for computing the centroids fails. Another common approach is simply to subtract the mean of the coordinates in each frame individually, i.e., neglecting the correction term in (6.4). As we learned above, however, this is similarly not optimal.

## 6.4 Closed-Form Solution for the 3D Coordinates

In the previous section, we obtained the closed-form solution for the translation parameter  $\mathbf{t}^i$ . Without loss of generality and to guarantee the optimal solution for the translations, this expression may now be substituted in the original Equations (6.1) and (6.2). Consequently, we get

$$C(\mathbf{m}) = \sum_i \sum_j (\mathbf{M}^i(\mathbf{x}_j - \bar{\mathbf{x}}^i) - (\mathbf{m}_j^i - \bar{\mathbf{m}}^i))^T (\mathbf{M}^i(\mathbf{x}_j - \bar{\mathbf{x}}^i) - (\mathbf{m}_j^i - \bar{\mathbf{m}}^i)) \delta_{ij}. \quad (6.5)$$

Let us define  $\mathbf{y}_j^i = \mathbf{x}_j - \bar{\mathbf{x}}^i$ . Now,

$$\frac{\partial C(\mathbf{m})}{\partial \mathbf{y}_j^i} = 2 (\mathbf{M}^i \mathbf{y}_j^i + (\mathbf{m}_j^i - \bar{\mathbf{m}}^i))^T \mathbf{M}^i \delta_{ij} \quad (6.6)$$

and

$$\frac{\partial \mathbf{y}_j^i}{\partial \mathbf{x}_j} = \frac{\partial}{\partial \mathbf{x}_j} \left( \mathbf{x}_j - \frac{\sum_{j'} \mathbf{x}_{j'} \delta_{ij'}}{\sum_{j'} \delta_{ij'}} \right) = 1 - \frac{\delta_{ij}}{\sum_{j'} \delta_{ij'}} = 1 - \frac{\delta_{ij}}{M^i}, \quad (6.7)$$

thus using the chain rule we get

$$\begin{aligned}\frac{\partial C(\mathbf{m})}{\partial \mathbf{x}_j} &= \sum_i \frac{\partial C(\mathbf{m})}{\partial \mathbf{y}_j^i} \frac{\partial \mathbf{y}_j^i}{\partial \mathbf{x}_j} \\ &= 2 \sum_i \left(1 - \frac{1}{M^i}\right) (\mathbf{M}^i(\mathbf{x}_j - \bar{\mathbf{x}}^i) - (\mathbf{m}_j^i - \bar{\mathbf{m}}^i))^T \mathbf{M}^i \delta_{ij}.\end{aligned}\tag{6.8}$$

Setting the derivative to zero, it follows

$$\begin{aligned}\frac{\partial C(\mathbf{m})}{\partial \mathbf{x}_j} &= \mathbf{0} \\ \Leftrightarrow \sum_i \left(1 - \frac{1}{M^i}\right) (\mathbf{M}^{iT} \mathbf{M}^i(\mathbf{x}_j - \bar{\mathbf{x}}^i) - \mathbf{M}^{iT}(\mathbf{m}_j^i - \bar{\mathbf{m}}^i)) \delta_{ij} &= \mathbf{0} \\ \Leftrightarrow \sum_i \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT} \mathbf{M}^i(\mathbf{x}_j - \bar{\mathbf{x}}^i) \delta_{ij} &= \sum_i \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT}(\mathbf{m}_j^i - \bar{\mathbf{m}}^i) \delta_{ij}.\end{aligned}\tag{6.9}$$

Let us mark the right hand side of the previous equation by  $\mathbf{b}_j$  that is a  $3 \times 1$  vector. From the left hand side we get

$$\begin{aligned}\sum_i \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT} \mathbf{M}^i(\mathbf{x}_j - \bar{\mathbf{x}}^i) \delta_{ij} &= \sum_i \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT} \mathbf{M}^i \left(\mathbf{x}_j - \frac{\sum_{j'} \mathbf{x}_{j'} \delta_{ij'}}{M^i}\right) \delta_{ij} \\ &= \sum_i \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT} \mathbf{M}^i \mathbf{x}_j \delta_{ij} - \sum_i \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT} \mathbf{M}^i \frac{\sum_{j'} \mathbf{x}_{j'} \delta_{ij'}}{M^i} \delta_{ij} \\ &= \underbrace{\left(\sum_i \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT} \mathbf{M}^i \delta_{ij}\right)}_{\triangleq \mathbf{H}_j} \mathbf{x}_j + \\ &\quad - \sum_{j'} \underbrace{\left(\sum_i \frac{1}{M^i} \left(1 - \frac{1}{M^i}\right) \mathbf{M}^{iT} \mathbf{M}^i \delta_{ij'} \delta_{ij}\right)}_{\triangleq \mathbf{G}_{j,j'}} \mathbf{x}_{j'} \\ &= \mathbf{H}_j \mathbf{x}_j - \sum_{j'} \mathbf{G}_{j,j'} \mathbf{x}_{j'},\end{aligned}\tag{6.10}$$

where  $\mathbf{H}_j$  and  $\mathbf{G}_{j,j'}$  are  $3 \times 3$  matrices. We may now write (6.9) and (6.10) in matrix form, i.e.,

$$\begin{pmatrix} \mathbf{H}_1 - \mathbf{G}_{1,1} & -\mathbf{G}_{1,2} & -\mathbf{G}_{1,3} & \cdots \\ -\mathbf{G}_{2,1} & \mathbf{H}_2 - \mathbf{G}_{2,2} & -\mathbf{G}_{2,3} & \cdots \\ -\mathbf{G}_{3,1} & -\mathbf{G}_{3,2} & \mathbf{H}_3 - \mathbf{G}_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ \vdots \end{pmatrix}. \quad (6.11)$$

We can not yet compute  $\mathbf{x}_j$ s directly as the coefficient matrix is singular. This is because we may choose the origin of the 3D coordinates arbitrarily; thus, there would be an infinite number of feasible solutions. Let us fix the 3D origin to the center of the all the 3D coordinates, i.e.,

$$\sum_j \mathbf{x}_j = \mathbf{0} \Leftrightarrow \mathbf{x}_1 = -\sum_{j \neq 1} \mathbf{x}_j. \quad (6.12)$$

Because  $\mathbf{A}\mathbf{x}_1 = -\mathbf{A}\sum_{j \neq 1} \mathbf{x}_j = -\sum_{j \neq 1} \mathbf{A}\mathbf{x}_j$  for any  $3 \times 3$  matrix  $\mathbf{A}$ , the above constraint may be encoded in the matrix equation (6.11) by removing and summing the negative first (matrix) column to the other columns. We get

$$\begin{pmatrix} -\mathbf{G}_{1,2} + \mathbf{G}_{1,1} - \mathbf{H}_1 & -\mathbf{G}_{1,3} + \mathbf{G}_{1,1} - \mathbf{H}_1 & \cdots \\ \mathbf{H}_2 - \mathbf{G}_{2,2} + \mathbf{G}_{2,1} & -\mathbf{G}_{2,3} + \mathbf{G}_{2,1} & \cdots \\ -\mathbf{G}_{3,2} + \mathbf{G}_{3,1} & \mathbf{H}_3 - \mathbf{G}_{3,3} + \mathbf{G}_{3,1} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ \vdots \end{pmatrix}. \quad (6.13)$$

Now, the coefficient matrix has one more row than columns. The equations are, however, redundant because using the chain rule and (6.12), we may write

$$\frac{\partial C(\mathbf{m})}{\partial \mathbf{x}_1} = \sum_{j \neq 1} \frac{\partial C(\mathbf{m})}{\partial \mathbf{x}_j} \frac{\partial \mathbf{x}_j}{\partial \mathbf{x}_1} = \sum_{j \neq 1} \frac{\partial C(\mathbf{m})}{\partial \mathbf{x}_j} (-1) = -\sum_{j \neq 1} \frac{\partial C(\mathbf{m})}{\partial \mathbf{x}_j}. \quad (6.14)$$

In the minimum, the right hand side vanishes; hence the first (or any other) equation is unnecessary, and we can write

$$\begin{pmatrix} \mathbf{H}_2 - \mathbf{G}_{2,2} + \mathbf{G}_{2,1} & -\mathbf{G}_{2,3} + \mathbf{G}_{2,1} & \cdots \\ -\mathbf{G}_{3,2} + \mathbf{G}_{3,1} & \mathbf{H}_3 - \mathbf{G}_{3,3} + \mathbf{G}_{3,1} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{b}_2 \\ \mathbf{b}_3 \\ \vdots \end{pmatrix}. \quad (6.15)$$

Let us denote the coefficient matrix, the coordinate vector, and the right hand side of (6.15) with  $\mathbf{A}$ ,  $\mathbf{X}$ , and  $\mathbf{B}$ , respectively, thus

$$\mathbf{A}\mathbf{X} = \mathbf{B}. \quad (6.16)$$

We have thus obtained the following result.

**Proposition 6.4.1** *Let  $\mathbf{A}$ ,  $\mathbf{X}$ , and  $\mathbf{B}$  be defined as in (6.16). Let us assume general motion so that the points do not lie on a critical surface and there are enough visible points, i.e.,  $\mathbf{A}$  is not singular. Then, we obtain the least squares estimate for  $\mathbf{X}$  from*

$$\hat{\mathbf{X}} = \mathbf{A}^{-1}\mathbf{B}, \quad (6.17)$$

*whereas  $\hat{\mathbf{x}}_1$  and  $\hat{\mathbf{t}}^1, \dots, \hat{\mathbf{t}}^N$  can be thereafter computed from (6.12) and (6.3), respectively.*

## 6.5 Application Examples

In this section, possible applications for the derived solution are considered. As we note in Section 6.5.1, the solution gives the closed-form maximum likelihood affine triangulation under missing data. In addition, we propose two iterative algorithms for the maximum likelihood affine reconstruction (Section 6.5.2) and a method for motion degeneracy identification (Section 6.5.3). Finally, in Section 6.5.4, we show how the result can be used for image registration in electron tomography, which is an example of the bundle adjustment under constrained motion.

### 6.5.1 Affine Triangulation

In general, triangulation involves computing the scene structure, given cameras and the corresponding image points. It is assumed that the measurement errors are only in points rather than in the projection matrices. However, naive triangulation by back-projecting the measured points fails because the back projected rays are skew in general. The optimal way is rather to estimate both the points in 3-space and their noise free projections by minimizing the square sum of the distances between the measured and estimated image points.

Hartley and Strum (1997) showed that, when using the full projective camera model, the maximum likelihood triangulation of two views can be non-iteratively computed by solving a sixth-order polynomial. Our result, on the other hand, derived in the previous section, gives a non-iterative, closed-form solution for ML affine triangulation of  $N$ -images with missing data. After computing the structure, from (6.17) and (6.12), and the translation vectors (6.3), the noise free point estimates can be directly computed using the model given in (6.1).

The reader may finally ask the following question: “Why should a triangulation for one point be dependent on the triangulation for all other points when the cameras are known? With given cameras, triangulation should be done separately for every point.” The answer is that we assume that we know only the inhomogeneous projection matrices; i.e., the translation part of the full homogeneous projection

matrix is assumed unknown. Under missing data, the translation part depends on all the points through the reconstructed 3D coordinates of the points according to (6.4). Therefore, we actually have more than a closed-form triangulation since our assumptions are weaker than known camera projection matrices.

### 6.5.2 Affine Reconstruction

The factorization algorithm for affine reconstruction (Tomasi and Kanade 1992) requires a measurement of each point in all views. This is clearly a major limitation since the matched points may be absent in certain views. In this section, using the result derived above, we propose two iterative algorithms for affine reconstruction where this limitation is eliminated. We search for the maximum likelihood estimate by fixed point algorithms, where missing data are iteratively updated by the most likely values given the current estimates for the parameters.

The proposed methods can be used in computing the affine multiple view tensors, i.e. affine fundamental matrix, affine trifocal tensor, and affine quadrifocal tensor, in an optimal way, assuming Gaussian noise model and global convergence of the algorithms. As stated before, no single feature point is required to be found in each image even though all the  $N$  images may be used simultaneously. The tensors can be computed directly from the camera projection matrices (see e.g. Hartley and Zisserman 2000).

#### Iterative Factorization

In the first step of the first proposed approach, one needs initial estimates for the projection matrices  $\mathbf{M}^i$ . To get these estimates, one may use some a priori information of the image positions or solve the parameters using smaller image sets with the factorization algorithm. Another possibility is Jacobs's (2001) method, which additionally estimates the translations in contrast to the earlier version (Jacobs 1997).

In the second step, the closed-form ML estimates for  $\mathbf{x}_j$  given  $\hat{\mathbf{M}}^i$ ,  $i = 1, \dots, N$ , are solved from (6.17) and (6.12). For the feature points  $j$  missing in image  $i$  ( $\delta_{ij} = 0$ ) we set

$$\mathbf{m}_j^i = \hat{\mathbf{M}}^i \hat{\mathbf{x}}_j + \hat{\mathbf{t}}^i, \quad (6.18)$$

where the estimates for  $\mathbf{t}^i$  are computed from (6.3). Now we may form the measurement matrix

$$\mathbf{W} = \begin{pmatrix} \mathbf{m}_1^1 - \hat{\mathbf{t}}^1 & \dots & \mathbf{m}_M^1 - \hat{\mathbf{t}}^1 \\ \vdots & \ddots & \vdots \\ \mathbf{m}_1^N - \hat{\mathbf{t}}^N & \dots & \mathbf{m}_M^N - \hat{\mathbf{t}}^N \end{pmatrix}. \quad (6.19)$$

In the third step, we seek the maximum likelihood solution for the projection matrices  $\mathbf{M}^i$ , given the measurement matrix  $\mathbf{W}$ . This is achieved by using the factorization method (Tomasi and Kanade 1992). Hence, we search for the matrix  $\hat{\mathbf{W}}$ , which minimizes the equation

$$\|\hat{\mathbf{W}} - \mathbf{W}\|_F, \quad (6.20)$$

where  $\|\cdot\|_F$  is the Frobenius norm, and can be decomposed as

$$\hat{\mathbf{W}} = \begin{pmatrix} \mathbf{M}^1 \\ \mathbf{M}^2 \\ \vdots \\ \mathbf{M}^N \end{pmatrix} (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_M). \quad (6.21)$$

Since the rank of  $\hat{\mathbf{W}}$  should be three, the solution which minimizes the Frobenius norm may be obtained by computing the singular value decomposition (SVD) for  $\mathbf{W} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ , and truncating the singular values by setting all the singular values to zero except the three largest. Equivalently, we get

$$\hat{\mathbf{W}} = \mathbf{U}'\mathbf{S}'\mathbf{V}'^T, \quad (6.22)$$

where  $\mathbf{U}'$  the matrix of the first three columns of  $\mathbf{U}$ ,  $\mathbf{S}'$  is the  $3 \times 3$  diagonal matrix of the three largest singular values, and  $\mathbf{V}'$  consists of the first three columns of  $\mathbf{V}$ . The maximum likelihood solution for  $\mathbf{M}$  is obtained, up to an affine transformation, by setting  $\hat{\mathbf{M}} = \mathbf{U}'\mathbf{S}'$ .

The above procedure may now be iterated until convergence, or in other words, until the Frobenius norm of the change in  $\hat{\mathbf{W}}$  becomes smaller than a predefined tolerance value. There is no guarantee of global convergence in general, but the essence of this algorithm is that *only* the missing data are iteratively replaced by the expected values in the second step, and everything else is computed in closed form. Therefore, the optimal solution is guaranteed when we do not have missing data and, consequently, the algorithm is robust to increasing amounts of missing data with any choice of initial estimate for the projection matrix. This can also be seen from the experimental results reported in Section 6.6. The algorithm is summarized in Algorithm 5.

### Bilinear Algorithm

An alternative algorithm for affine reconstruction is obtained if we do not regard the missing feature points in the image set as the missing data *in the algorithm point of view*, but instead all the 3D coordinates of the feature points. Then, we obtain a bilinear algorithm similar to Morris and Kanade's (1998) with the difference that our algorithm is able to handle missing feature points in an optimal way (see Section 6.3 for explanation).

---

**Algorithm 5** Iterative Factorization (IF) for Affine Reconstruction with Missing Data

---

1. Set initial values for projection matrices  $\mathbf{M}^i$ , e.g., by Jacobs's (2001) method.
  2. Compute the closed-form solution for the feature point 3D coordinates  $\mathbf{x}_j$ , given  $\hat{\mathbf{M}}^i$ ,  $i = 1, \dots, N$  from (6.17), and (6.12). Compute new estimates for  $\mathbf{t}^i$  from (6.3). Use the model  $\mathbf{m}_j^i = \mathbf{M}^i \mathbf{x}_j + \mathbf{t}^i$  to replace the absent feature point coordinates in the measurement matrix  $\mathbf{W}$ .
  3. Use the factorization algorithm to obtain the ML estimates for  $\mathbf{M}$ , i.e., compute the singular value decomposition for the measurement matrix  $\mathbf{W} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ .  $\hat{\mathbf{M}}$  is obtained, up to an affine transformation, by multiplying the first three columns of  $\mathbf{U}$  by the first three singular values in  $\mathbf{S}$ , respectively. Repeat Steps 2 and 3 until convergence.
- 

The bilinear algorithm is in fact quite similar to the iterative factorization algorithm. We similarly need initial estimates for the projection matrices  $\mathbf{M}^i$ . The second step computes ML estimates for the reconstructions, given  $\mathbf{M}^i$ , using (6.17) and (6.12), and now these values are used as the “missing data” in the third step.

In the third step, new estimates for  $\mathbf{M}^i$ , given  $\hat{\mathbf{x}}_j$ ,  $j = 1, \dots, M$ , are computed. By differentiating (6.5) with respect to  $\mathbf{M}^i$  and setting the result zero, it follows

$$\mathbf{M}^i \sum_j (\mathbf{x}_j - \bar{\mathbf{x}}^i)(\mathbf{x}_j - \bar{\mathbf{x}}^i)^T \delta_{ij} = \sum_j (\mathbf{m}_j^i - \bar{\mathbf{m}}^i)(\mathbf{x}_j - \bar{\mathbf{x}}^i)^T \delta_{ij}. \quad (6.23)$$

Assuming that the vectors  $\mathbf{x}_j - \bar{\mathbf{x}}^i$ , visible in image  $i$ , span the 3-space, we get the ML estimate for  $\mathbf{M}^i$ , given  $\mathbf{x}_j$ ,  $j = 1, \dots, M$  from

$$\begin{aligned} \hat{\mathbf{M}}^i &= \left( \sum_j (\mathbf{m}_j^i - \bar{\mathbf{m}}^i)(\mathbf{x}_j - \bar{\mathbf{x}}^i)^T \delta_{ij} \right) \left( \sum_j (\mathbf{x}_j - \bar{\mathbf{x}}^i)(\mathbf{x}_j - \bar{\mathbf{x}}^i)^T \delta_{ij} \right)^{-1} \\ &= \hat{\mathbf{C}}_{\mathbf{m},\mathbf{x}}^i \left( \hat{\mathbf{C}}_{\mathbf{x}}^i \right)^{-1}, \end{aligned} \quad (6.24)$$

where  $\hat{\mathbf{C}}_{\mathbf{m},\mathbf{x}}^i$  is the sample cross covariance matrix between  $\mathbf{m}_j^i$  and  $\mathbf{x}_j$ , visible in image  $i$ , and  $\hat{\mathbf{C}}_{\mathbf{x}}^i$  is the sample covariance matrix of  $\mathbf{x}_j$ , visible in image  $i$ . The translations are handled invisibly and are finally obtained from (6.3). The second algorithm is summarized in Algorithm 6.

---

**Algorithm 6** Bilinear Algorithm (BA) for Affine Reconstruction with Missing Data

---

1. Set initial values for inhomogeneous projection matrices  $\mathbf{M}^i$ , e.g., by Jacobs's (2001) method.
  2. Compute the closed-form solution for the feature point 3D coordinates  $\mathbf{x}_j$  given  $\hat{\mathbf{x}}_j$ ,  $i = 1, \dots, N$  from (6.17), and (6.12).
  3. Use (6.24) to get the ML estimates for  $\mathbf{M}^i$ , given  $\hat{\mathbf{x}}_j$  and  $j = 1, \dots, M$ . Repeat Steps 2 and 3 until convergence. Estimates for  $\mathbf{t}^i$  may be finally computed from (6.3).
- 

### 6.5.3 Degeneracy Identification

In the previous section, we obtained the ML solution for affine reconstruction assuming general motion, and convergence to the global minimum. It is, however, possible that the motion is degenerate. For example, when the cameras are moved along a single line and directed in the same direction, it is not possible to solve the depth from the affine projections. An even simpler degenerate case is that when the cameras are not moved at all, or when the displacement is very small. Then, the depth calculation also confronts an ill-posed problem.

In the degenerate cases, the information obtained from the measured coordinate values may not provide a unique solution, or the solution may be poorly conditioned and highly sensitive to noise. Most degenerate conditions can be observed by examining the matrix  $\mathbf{A}$  in (6.16). In fact, if there were ambiguity in the 3D coordinates, in addition to the already fixed origin, the determinant of  $\mathbf{A}$  would approach zero, and the inverse would not exist.

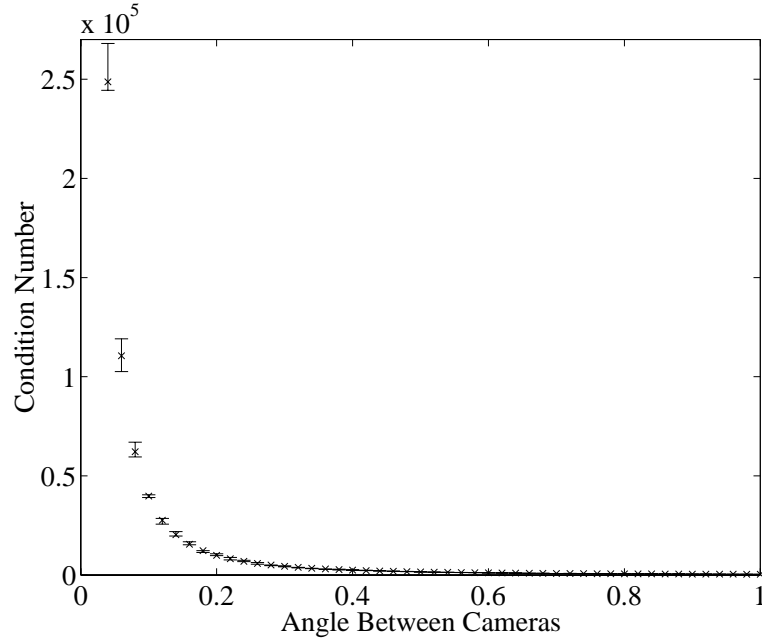
The determinant value itself is not a good measure of ill-conditioning even though the vanishing determinant is equivalent to singularity (Golub and Loan 1996). A far better measure for degeneracy is the condition number

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|, \quad (6.25)$$

where, for example, the 2-norm may be used. The condition number approaches infinity for singular  $\mathbf{A}$ . Geometrically the 2-norm condition measures the elongation of the hyper ellipsoid  $\{\mathbf{Ax} \mid \|\mathbf{x}\| = 1\}$  (Golub and Loan 1996).

In short, we therefore propose that ill-posed or highly uncertain cameras could be identified by examining the condition number of the coefficient matrix  $\mathbf{A}$  in (6.16). For instance, in either of the proposed algorithms for affine reconstruction, the condition number can be evaluated on each iteration step to indicate how reliably





**Figure 6.1:** Illustration of the condition number  $\kappa$  of the coefficient matrix  $\mathbf{A}$  near a degenerate configuration. Here, 200 points were randomly generated in a cube and the points were projected via orthographic cameras onto ten image planes, where Gaussian noise with  $\sigma = 1$  was added to the projections. In addition, 10% of the projections were randomly set absent. The cameras were placed along a line in space with a constant displacement between consecutive cameras. The cameras were rotated over the axis perpendicular to the line and the optical axis of the first camera such that the angular separation between consecutive cameras was kept constant. The  $x$ -axis of the graph indicates the angular separation where zero means a degenerate configuration. The condition number has been computed as the median over 100 trials and the first and third quartile have been used as the error bounds, respectively.

the 3D coordinates are determined. Figure 6.1 shows an example. Clearly, not only structure or motion degeneracies imply large condition number, but also inadequate visibility of the points. We regard this as an advantage since we now have an overall measure for the stability of the reconstructed points. Similarly, if the second, bilinear algorithm is selected, the sample cross correlation matrix  $\hat{\mathbf{C}}_{\mathbf{x}}^i$  in (6.24) is singular if and only if there are less than three linearly independent points, with respect to  $\bar{\mathbf{x}}^i$ , in the 3-space that are visible in image  $i$ . Condition number of these matrices may hence also be used to discover these kinds of degeneracies.

### 6.5.4 Image Registration in Electron Tomography

In axial electron tomography (see Chapters 7 and 8), the motion to be solved is constrained to rotation around one axis and its characterized by the model (Lawrence 1992) (compare to (6.1))

$$\mathbf{m}_j^i = s^i \mathbf{R}_\alpha^i \mathbf{P} \mathbf{R}_\beta^i \mathbf{x}_j + \mathbf{t}^i, \quad (6.26)$$

where  $s^i$  is a scaling factor,  $\mathbf{R}_\alpha^i$  is a  $2 \times 2$  rotation matrix,  $\mathbf{R}_\beta^i$  is  $3 \times 3$  rotation matrix describing the tilting operation around the  $y$ -axis,  $\mathbf{t}^i$  is a translation vector for image  $i$ , and  $\mathbf{P}$  is an orthographic projection matrix:

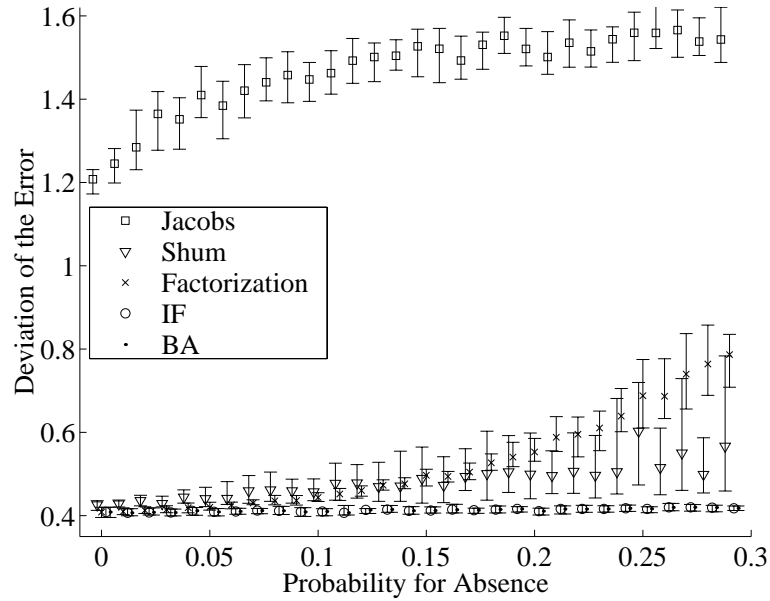
$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \quad (6.27)$$

The objective is to find optimal values for the parameters  $s^i$ ,  $\alpha^i$ ,  $\beta^i$ ,  $\mathbf{t}^i$ , and  $\mathbf{x}_j$ , given the realizations for the fiducial marker coordinates  $\mathbf{m}_j^i$ . This is a nonlinear optimization problem otherwise known as *bundle adjustment*, which can be solved using, for example, the conjugate gradient method (Press, Teukolsky, Vetterling and Flannery 1993). However, optimization requires initial guesses for parameters. For all other parameters, proper values may be addressed, but  $\mathbf{x}_j$  are problematic, as we assume that each marker is not found in every image. To solve the problem, we may now use (6.17) and (6.12), given all the other initial guesses.

On the other hand, it is further possible to formulate the optimization problem such that we search for the closed-form optimal values for  $\mathbf{x}_j$  (and  $\mathbf{t}^i$ ) on each iteration step in the optimization. Then, the number of parameters to be optimized does not depend on the number of markers used, but only on the number of images! This allows using a huge number of markers because they do not affect the direct memory use of the optimization routine. As a disadvantage, however, the inversion needed for computing the closed-form solution tends to take more computation time than is saved by reducing the dimensionality of the parameter space.

## 6.6 Experiments

In this section, we present experiments with the proposed two iterative algorithms for affine reconstruction. We have compared these methods to Jacobs's (2001) and Shum et al.'s (1995) methods. Because Shum's method assumes known translations, the translations given by the Jacobs' method are used there. In addition, the result of Jacobs' method is also used as the initial guess for Shum's and the proposed methods. With synthetic data, we also compare the methods



**Figure 6.2:** Affine reconstruction with synthetic data and missing measurements. The  $x$ -axis gives the probability for any projection to be absent in an image, but to allow comparison with the standard factorization, at least 4 points were set visible in all views. The  $y$ -axis gives the standard deviation of the Euclidean distance between the true projections and the estimated noise free ones in an independent test set. The experiment was repeated 50 times, and the median and the first and third quartiles of the scores were used as the result and the error bounds, respectively. The results are slightly shifted with respect to one another to facilitate inspection.

to factorization (Tomasi and Kanade 1992), which that is computed using only the points that are visible in all views.

The synthetic experiments were performed as follows. 200 points were randomly generated within a cube. The points were projected via orthographic projection onto ten image planes. To the projections Gaussian noise was added with  $\sigma = 1$ . Each projection was randomly set absent with a constant probability during each set of trials. In addition, at least 4 points were set to be visible in all views to allow comparison with the standard factorization method. We made 50 trials for each probability value and generated independent point sets for evaluation. An individual error score was computed as the standard deviation of the distance between the true noise free projections and the estimated noise free ones, obtained using standard triangulation (translations are now given). The median of the scores was used as the result and the error bounds were correspondingly set to the first and third quartile. The results are shown in Figure 6.2.

Jacobs's method can be considered relatively stable to the increasing amount of missing data, and its expected bias can be clearly observed. There is presumably

**Table 6.1:** Results for the affine reconstruction with the four methods and three real image sets. The score is the standard deviation of the distance between the measured coordinates and estimated noise free counter parts, i.e., the reprojection error.

Image Set	Jacobs	Shum	IF	BA
Mitochondrion	5.7	0.94	0.78	0.78
Microvillus	3.1	1.4	0.82	0.82
Section	0.94	0.54	0.46	0.46

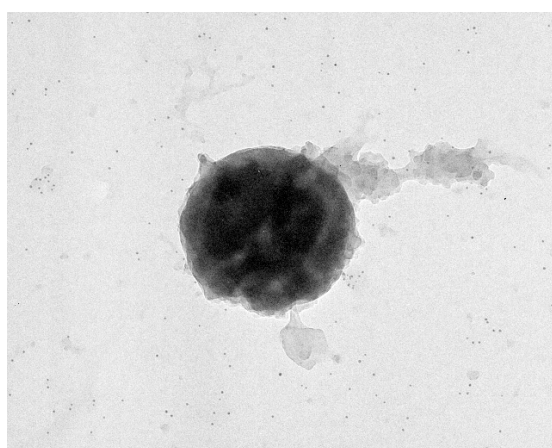
an additional bias in its translation estimates, which consequently deteriorates the results of Shum’s method. As expected, when the probability of losing a point in an image increases, the factorization algorithm has fewer data and the result gets worse. The proposed iterative algorithms give practically uniform results, as they should. Their error grows slower than factorization because the ML estimates are computed using all the observed data.

As real data, we used three sets of electron microscope images of a mitochondrion, a microvillus, and a section of another mitochondrion (Fig. 6.3). In the case of mitochondrion, the tracked features were gold beads that can be automatically found in the images, as proposed in Chapter 7, while the tracks are relatively long. There were 40 images in the set, and only tracks longer than 14 were used. The measurement matrix consisted of 166 tracks in total and approximately 35% missing data. Of all the tracks, 10 covered the entire set.

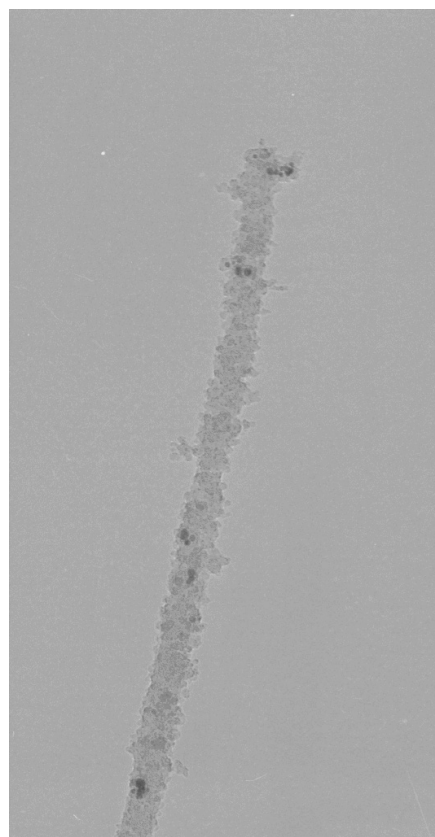
There were 10 images in the other set (microvillus). In that set, corner features, obtained by the improved Harris detector (Schmid et al. 2000), were tracked; see Chapter 8. Only the tracks longer than 4 were used. The measurement matrix consisted of 49 tracks and approximately 43% missing data. One track covered the entire sequence but the median length of tracks was no longer than 5.

In the third set (mitochondrion section), corner features were also tracked. Only the tracks longer than 4 were used when the measurement matrix consisted of 46% missing data. There were 69 tracks, none of which covered the entire sequence. The median length of the tracks was again 5 here. Table 6.1 shows the results for all the test sets.

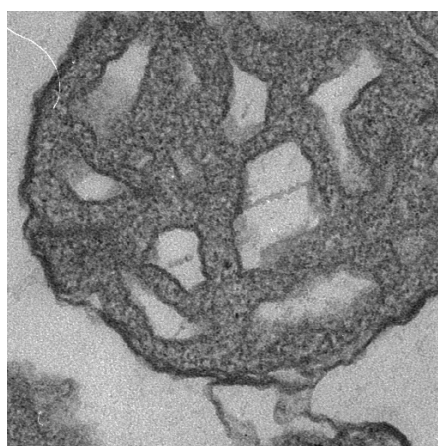
The results for real data are very much what can be expected. Since these image sets have been obtained tilting over an axis, as presented in Section 6.5.4, we have additionally computed the error with bundle adjustment using the constrained motion model (6.26). The bundle adjustment resulted in deviations of 0.89, 0.87, and 0.47 for the mitochondrion, microvillus, and section respectively. Furthermore, these results were in good agreement with the results in Table 6.1.



(a)

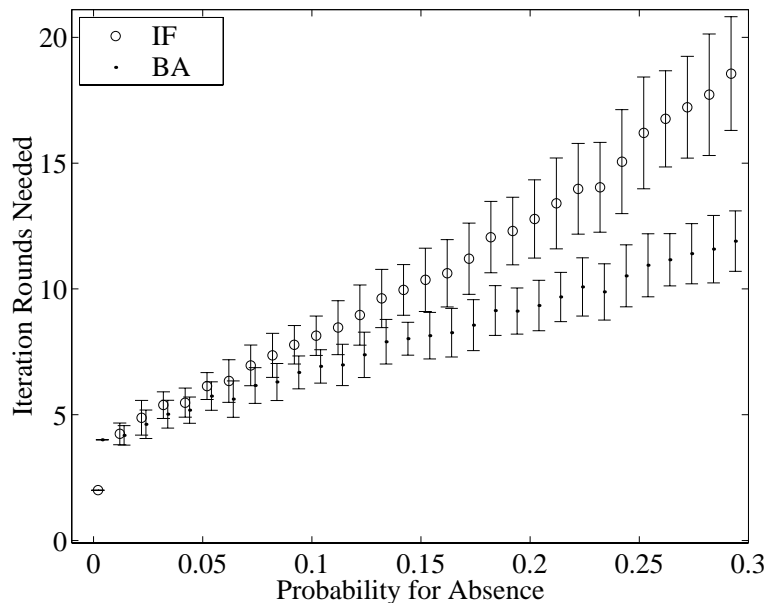


(b)



(c)

**Figure 6.3:** Example images from the real image sets: (a) mitochondrion; (b) microvillus; and (c) section of a mitochondrion.



**Figure 6.4:** Iteration steps needed for convergence in the two distinct iterative algorithms, corresponding to the experiment in Fig. 6.2. The mean over 50 trials were used as the result, and the sample standard deviations as error bounds, respectively. The amount of iteration steps needed in the iterative factorization (IF) increases with increasing amounts of missing data compared to the bilinear algorithm (BA). The square sum of the change in  $\mathbf{W}$  with the tolerance value  $10^{-7}$  was used as the convergence measure.

## 6.7 Note on Algorithmic Complexity

The closed-form solution for the reconstruction of  $M$  points in (6.17) involves the inversion of a  $(3M - 3) \times (3M - 3)$  matrix, which is relatively costly. Nevertheless, the complexity of a matrix inversion is not larger than matrix multiplication because the inversion and multiplication algorithms are equivalent. For example, Strassen's algorithm for  $n \times n$  matrix runs in  $O(n^{2.81})$  Time, and the best known algorithm, though less practical, has the upper bound of  $O(n^{2.376})$  (Cormen, Leiserson and Rivest 1996). In practice, it has been reported that Strassen's algorithm starts to be advantageous compared to the basic  $O(n^3)$  algorithm when  $n > 100$  (Cohen and Roth 1976). Hence in our case, Strassen's algorithm for matrix inversion should be favored if  $M > 30$ , approximately.

In the first proposed algorithm for affine reconstruction, the maximum likelihood estimates for inhomogeneous projection matrices are computed using singular value decomposition in the third step. The complexity of factorization for  $2N \times M$  measurement matrix is  $O(NM^2)$ . On the other hand, the third step in the bilinear algorithm algorithm is less complex since it runs in  $O(NM)$  time.

When the complexity of the two distinct algorithms is compared, however, the number of iteration rounds needed for convergence is crucial. The experiments with synthetic data suggests, see Fig. 6.4, that with small amount of missing data, in addition to being very reliable, the iterative factorization converges fast. Nevertheless, the bilinear algorithm seems to converge faster if there are large amounts of missing data.

## 6.8 Summary

Observing motion in  $N$  images under affine projection, we have derived the closed-form solution for the translation vectors and 3D coordinates of the tracked feature points, given the inhomogeneous projection matrices in each image, but no feature point is required to be found in every image. The derived solution implies closed-form solution for the optimal affine triangulation, and two iterative algorithms for affine reconstruction. The approaches are iterative generalizations of the well known affine factorization algorithm for missing data, and give the maximum likelihood solution for the structure and affine projection matrices, and thereby also for the affine multiple view tensors, given all the tracked feature points. Other applications of the closed-form solution were shown to be in degenerate motion identification as well as in bundle adjustment of constrained motion such as electron microscope image registration for electron tomography. The performed experiments accord with the theoretical considerations.





## Chapter 7

# TEM Image Alignment with Fiducial Markers<sup>1</sup>

The last two chapters of this thesis consider how the methods of the previous chapters can be used in aligning transmission electron microscope (TEM) images for electron tomography; registration is essential to performing the 3D reconstruction of a TEM tilt-series. So far, the problem is solved by either manually showing the corresponding fiducial markers from the set of images, or automatically using simple correlation between the images on several rotations and scales. The present solutions, however, share the problem of being inefficient and/or inaccurate. In this chapter, we propose a method in which the registration is automated using conventional colloidal gold particles as reference markers between images. The alignment problem is divided into several subproblems: (1) finding initial matches from successive images, (2) estimating the epipolar geometry between consecutive images, (3) finding and localizing the gold particles with subpixel accuracy in each image, (4) predicting the probable matching gold particles using the epipolar constraint and its uncertainty, (5) matching and tracking the gold beads through the tilt series, and (6) optimizing the transformation parameters for the whole image set. The results show not only the reliability of the suggested method but also a high level of accuracy in alignment, since practically all the visible gold markers can be used.

---

<sup>1</sup>The published conference papers regarding this chapter are Brandt and Heikkonen (2000*a*) and Brandt, Heikkonen and Engelhardt (2001*a*); the journal version is Brandt, Heikkonen and Engelhardt (2001*c*).

## 7.1 Introduction

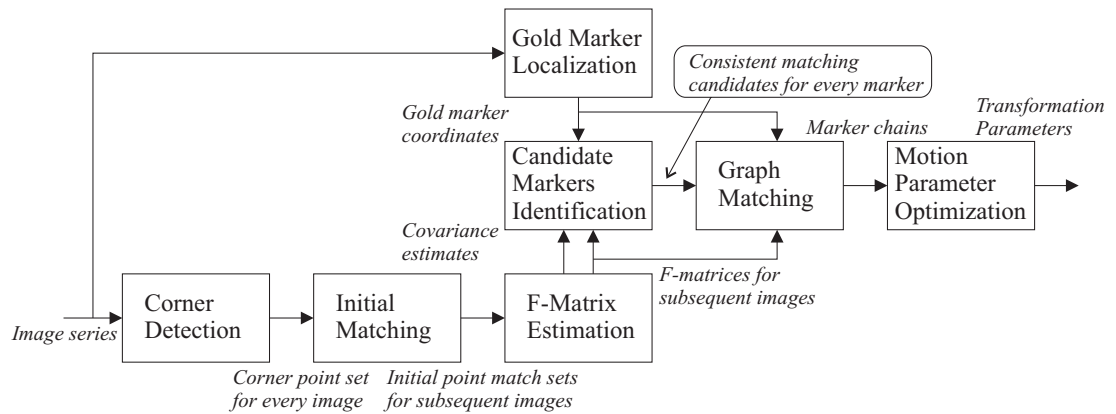
In structural biology, electron tomography (ET) is used to reconstruct three-dimensional objects such as macromolecules, viruses, and cellular organelles to learn their three-dimensional structures and properties. The reconstruction is made from a set of transmission electron microscope (TEM) images, which may be obtained by tilting the specimen stage by small angular increments (single axis tilting) or a fixed angle and rotating by constant increments in the specimen plane (conical tilting) (Frank 1992). In this work, we use a set of two-dimensional electron microscope images of a three-dimensional object that are obtained by tilting the specimen holder around a single axis, as described in detail in Engelhardt (2000).

Traditionally, the alignment is solved by either manually showing the corresponding markers from the set of images or automatically using simple correlation between the images on several rotations and scales. The manual approach has the disadvantage of being time-consuming for the user because, at minimum, three (four in unconstrained motion; Ullman 1979) identical markers must be shown in all images, but in practice to obtain accurate results, one should show many more. The previous correlation-based (Guckenberger 1982, Frank and McEwen 1992) automatic methods are, however, much more inaccurate because the transformations are computed between the consecutive images of the tilt series and therefore the errors accumulate along the image series.

The major advantage of using gold beads as fiducial markers is the fact that they can be localized accurately, even with subpixel precision. Independent of our work, there have been some attempts to automate the gold marker collecting and matching process (Fung, Liu, Ruijter, Chen, Abbey, Sedat and Agard 1996, Ress, Harlow, Schwarz, Marshall and McMahan 1999), but these approaches require almost ideal imaging conditions to work without user intervention and, in addition, they unnecessarily carry the burden of the previous, naive, cross-correlation-based registration methods. The purpose of this work is therefore to use some recent advances from the computer vision field to find a solution to the alignment problem.

The image series used in this work consists of 41 images that are taken onto a film by tilting the specimen stage by  $3^\circ$  increments from  $-60^\circ$  to  $60^\circ$  (Engelhardt 2000). The images are thereafter digitized from the film by a scanner, which causes some random rotation and shift to the images. As a result, if the images were shown as a movie, the motion would not be smooth and predictable. This makes our approach even more general than, e.g., the approach of Fung et al. (1996), because the prior information of the motion smoothness may not be used.

We have solved the alignment problem by dividing it into subproblems as Fig. 7.1 illustrates. In the first stage, the initial correspondence between successive im-



**Figure 7.1:** Flow chart of the proposed method. Input and output values for each phase are displayed in italics.

ages, based on detected corner points using correlation and relaxation techniques, and affine epipolar geometry are estimated, as described in Chapter 3. The epipolar constraint is needed in tracking the gold markers in the tilt series since it reduces the number of possible matches significantly. Before the final marker matching and tracking (Section 7.3), the gold beads must be localized as described in the following section. The transformation parameter optimization is presented in Section 7.4, and the viability of the proposed method is demonstrated in Section 7.5.

## 7.2 Fiducial Marker Localization

After the epipolar geometry has been estimated and before we can use it in matching and tracking the gold beads, the gold markers must be localized in the images. The major advantage of using fiducial gold markers as the tracked image features in the registration problem is that they can be reliably localized even with subpixel precision. To automatically search for the markers, we use the a priori information that the markers are spherical and that their size does not differ much from one to another. This allows us to represent a marker image by a filled circle. To model the effect of the point-spread function caused by the imaging process, the synthetic prototype is filtered by a Gaussian kernel.

We continue by computing the normalized correlation between the synthetic template and the original image (Fig. 7.2a and b). The correlation has the largest values on the most probable marker locations which can thus be obtained by thresholding on a certain value. However, spurious responses are also possible from areas that do not correspond to any marker. These false positions can be

discarded by using the shape information of the fiducial gold markers, as described below.

It can be observed that the markers also have circle-shaped responses in the thresholded correlation image (Fig. 7.2c). Therefore we could directly compute the circularity of these areas in order to determine whether a thresholded area corresponds to a marker. Here circularity is defined as (Han, Lee and Hwang 1994, Brandt 1999)

$$\zeta = \frac{4\pi A}{l^2}, \quad (7.1)$$

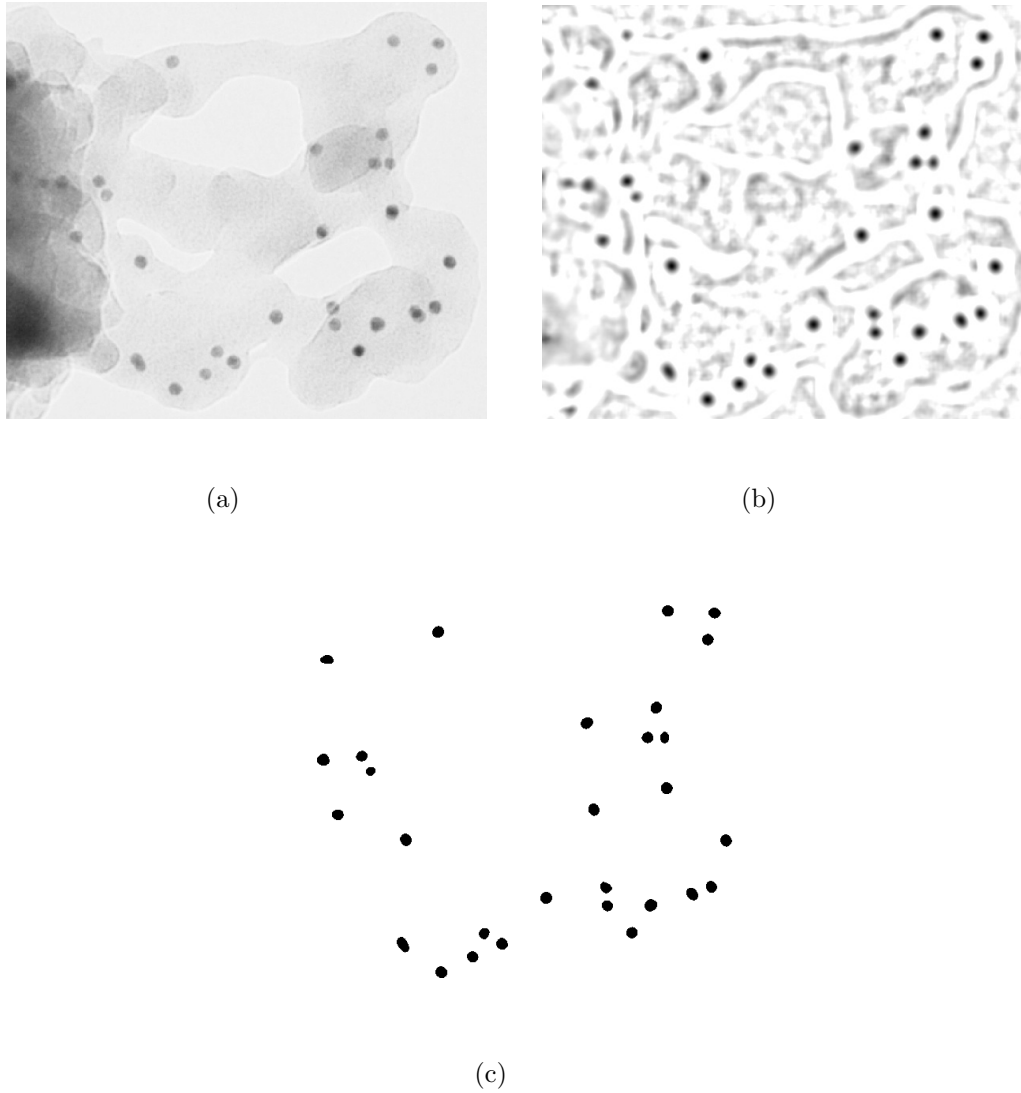
where  $l$  is the length of the perimeter of the region and  $A$  is the corresponding area. A circle then obtains roundness of unity and other shapes less than 1. There are fewer spurious responses if the threshold value is high—then, however, the high-correlation areas would be relatively small and the circularity would not give a good measure of the shape. To cope with this we suggest using similar thresholding with hysteresis as used in the Canny (1986) edge detector. That is, first the correlation image is thresholded with a high threshold value but also those regions that are above a lower threshold and connected to the areas of high correlation are considered.

The resulting binary image is processed by computing circularity for each separate region (in the four-neighborhood sense), which allows the classification according to the circularity score. The discrete area  $A$  in (7.1) is computed by weighting region elements in the way proposed in Pratt (1991) in order to compensate for the effect of discretation. The boundary length  $L$  is here regarded as the eight-connected outer boundary and its length is calculated as its *area* similarly weighted as above. The circularity threshold was set to a value of 0.8. The centers of mass of the circular classified areas are thereafter computed, after which the correlation maximum location is refined. This can be done with subpixel precision using the standard template matching technique.

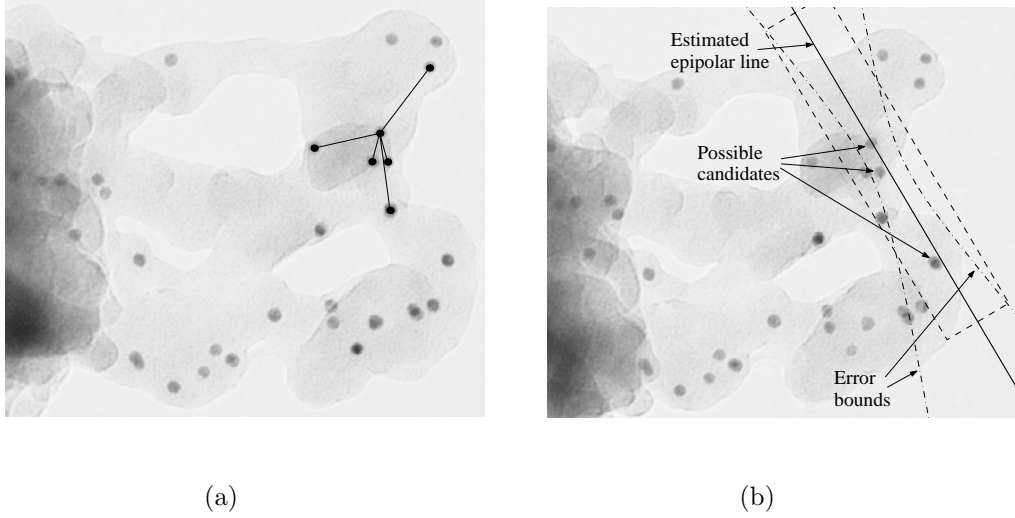
The proposed procedure reliably detects markers on image regions where even a direct thresholding of the intensity would not give satisfactory results. As a matter of fact, the markers are found even from areas where the task would be difficult for humans due to low contrast and noise. In addition, because the shape information is used, such markers that overlap in the image may also be discarded. The eventual overlaps are not taken into account in Ress et al. (1999) which may consequently induce a failure in the marker tracking.

### 7.3 Marker Matching and Tracking

When the gold beads are found, we solve the correspondence between the markers in consecutive images. In order to match markers in two images, some neighborhood information must be used because the beads look similar to one another.



**Figure 7.2:** (a) Fiducial gold markers in the image. (b) Correlation image computed with the template prototype; the darkest areas indicate the strongest correlation. (c) Correlation image after the thresholding with hysteresis.



**Figure 7.3:** (a) A graph corresponding to a marker and its five nearest neighbors. (b) Search region for the marker in the second image.

We propose that the neighborhood information could be taken into account by regarding the task as a graph matching problem where the correspondence between graphs, composed of markers and their neighbors, is compared (see Fig. 7.3a). The matching process is simplified by using the epipolar geometry estimated in Chapter 3, which, as we will see, will serve as an efficient constraint in selecting matching candidates.

### 7.3.1 Finding Candidates

In general, if the epipolar geometry was not used in the matching process we should consider every marker pair a possible match. We may, however, use the epipolar geometry to reduce the number of possible matches substantially. If the F-matrix was known perfectly, we might consider markers lying only on the epipolar lines. But because there is always some error in the F-matrix estimation, we examine the area inside the epipolar envelope (Csurka et al. 1997) or the error bounds of the epipolar lines (see Fig. 7.3b) which are determined by the F-matrix covariance matrix estimate (3.16).

The epipolar envelope is usually a hyperbola of two sheets on both sides of the estimated epipolar line. Sometimes the error bounds may also be far too pessimistic, so here we search only the intersection of the epipolar band and the rectangular area centered at the most probable point (see Section 4.2) with width and height of 250 and 50 pixels, respectively, as Fig. 7.3b indicates. The markers that are

found in this region in the second image are specified as candidate matches for the marker in the first image. Usually there are no more than a couple of possible candidates thus the use of the epipolar geometry reduces the possibility of a false match notably.

### 7.3.2 Graph Matching

The correspondence between markers in two images is identified by forming and comparing the neighborhood graphs. The graphs are here associated with the coordinates of the reference marker and its  $k$ -nearest neighbors. The graph matching procedure should be neither scale nor rotation invariant because we have the a priori knowledge that successive images do not differ much from one another. However, we should have robustness for both rotation and scale. In addition, the images may, however, have a shift in rotation due to the manual scanning of the film into a digital form (Engelhardt 2000).

To remove the effect of the ambiguity in rotation we first transform the graphs such that the epipolar lines are horizontal. As the affine camera model is used this is achieved by an affine transform of the graph where the angle can directly be determined from the fundamental matrix (Xu and Zhang 1996); the epipolar lines become horizontal by simple rotations of the graphs<sup>2</sup> by  $-\alpha$  in the left image where  $\alpha$  is given by

$$\alpha = \tan^{-1} \frac{f_{31}}{f_{32}}, \quad (7.2)$$

and by  $-\alpha'$  in the right image as

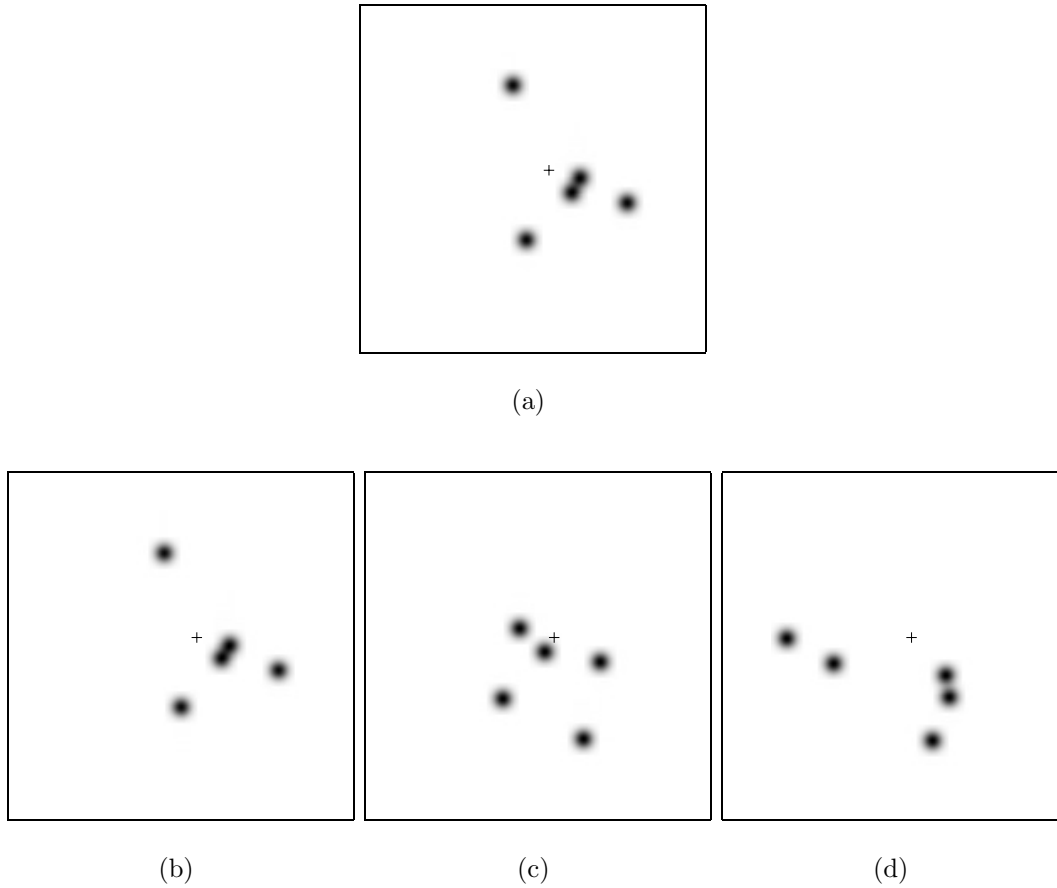
$$\alpha' = \tan^{-1} \frac{f_{13}}{f_{23}}. \quad (7.3)$$

The scale difference between images can be compensated for by scaling the graphs in the second image by the factor  $s$ , where

$$s = \sqrt{\frac{f_{13}^2 + f_{23}^2}{f_{31}^2 + f_{32}^2}}. \quad (7.4)$$

After the coordinate transform, for each graph a discrete 2D neighborhood uncertainty or impulse map is formed: In the reference marker centered coordinate system we set the values as 1 for the locations of the neighboring markers and 0 anywhere else. Between the images, the location uncertainties are modeled by a symmetric Gaussian distribution with deviation  $\sigma$  thus the maps are convolved with a two-dimensional Gaussian kernel (Fig. 7.4). A convenient choice for  $\sigma$  is the radius used for the synthetic template in Section 7.2 .

<sup>2</sup>The image origin is assumed to be in the left-hand corner of the image, y-axis downward.



**Figure 7.4:** Location uncertainty maps of the five nearest neighbors of the reference markers shown in Fig. 7.3. The reference marker locations are marked with a plus sign whereas the maps are obtained by transforming the reference marker centered coordinate frames so that the epipolar lines are horizontal and convolving the impulse map of the five nearest marker coordinates with a Gaussian kernel. The neighborhood of the reference marker in the first image (a) is compared for the neighborhoods of the candidate markers in the second image (b), (c), and (d). Clearly, there is correspondence between (a) and (b), which is observed by computing the normalized correlation between the uncertainty maps.



The purpose of the above maps is to encode the location information of the markers in a robust way while the neighboring information is preserved. The comparison of the graphs is performed by computing the normalized correlation between the uncertainty maps of the matching candidates where the maximum correlation scores exceeding 50% are taken as correspondences. In addition, we use 5 and 10 neighbors and combine the results since more matches can be found in this way compared to a single graph.

The above computations are finally made for all consecutive image pairs in a row and the results are saved in an appropriate data structure. The marker tracking is thereafter rather straightforward to implement since all that must be further found are the individual marker coordinates in each image and the image numbers where the chain starts and ends.

## 7.4 Parameter Optimization

The final stage in solving the alignment is the motion parameter optimization which is needed in transforming images to a common coordinate plane. As the imaging operation model we use the model defined by Lawrence (1992) with slight modifications. The  $j^{\text{th}}$  marker coordinates  $\mathbf{m}_j^i = (x_j^i \ y_j^i)^T$  in the  $i^{\text{th}}$  image are related to the corresponding 3D coordinates  $\mathbf{x}$  as

$$\mathbf{m}_j^i = s^i \mathbf{R}_\alpha^i \mathbf{P} \mathbf{R}_\beta^i \mathbf{x}_j + \mathbf{t}^i, \quad (7.5)$$

where  $s^i$  is a scaling factor,  $\mathbf{R}_\alpha^i$  is a  $2 \times 2$  rotation matrix associated with the angle  $\alpha$ ,  $\mathbf{R}_\beta^i$  is  $3 \times 3$  rotation matrix describing the tilting operation around the  $y$ -axis,  $\mathbf{t}^i$  is a translation vector for image  $i$ , and  $\mathbf{P}$  is an orthographic projection matrix:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \quad (7.6)$$

It is reasonable to assume that the measurement error, obtained by localizing the center of the gold markers from images, is normally distributed. ML estimates for the unknown transformation parameters are therefore obtained by minimizing the cost function (6.2). This form is thus a generalized version of that proposed by Lawrence (1992) because not all markers used are required to be found in all the images of the tilt series.

In order to obtain unique parameters in the minimization, a reference coordinate system must be set. Here, the rotation matrix  $\mathbf{R}_\beta^{i_r}$  is set to the identity matrix for the reference image  $i_r$  and the scale factor  $s^{i_r}$  is set to unity. In addition, we may arbitrarily choose the location of the origin; a convenient choice is the center

of the gold markers<sup>3</sup>, i.e.,

$$\sum_j \mathbf{x}_j = \mathbf{0}. \quad (7.7)$$

When the origin is chosen in this way the least-squares estimate for  $\mathbf{t}^i$  is obtained from (6.3), hence,

$$\hat{\mathbf{t}}^i = \frac{\sum_j \mathbf{m}_j^i \delta_{ij}}{\sum_j \delta_{ij}} - s^i \mathbf{R}_\alpha^i \mathbf{P} \mathbf{R}_\beta^i \frac{\sum_j \mathbf{x}_j \delta_{ij}}{\sum_j \delta_{ij}}, \quad (7.8)$$

which can now be substituted into (7.5) and (6.2).

We still need some initial guess for the parameters to be optimized. The scale parameters  $s^i$  can be initialized to unity and the rotation parameters  $\alpha^i$  to zero. If the tilt angles  $\beta^i$  are not known accurately enough, they can also be optimized as is done here. The initial values are directly obtained from the assumed orientation of the goniometer. However, we have noticed that if the number of markers is small, the  $\beta^i$  angles should not be optimized, but their assumed values should be used. The explanation for this is clear: if there are lots of parameters but few observations, the model is too complex or flexible. Then the result is that the model *overfits* to the data, or more accurately, to the noise (see e.g. Cherkassky and Mulier 1998). Overfitting can be avoided by restricting the complexity of the model or here by fixing  $\beta^i$ . The initial estimates for  $\mathbf{x}_j$  were computed as proposed in Section 6.5.4.

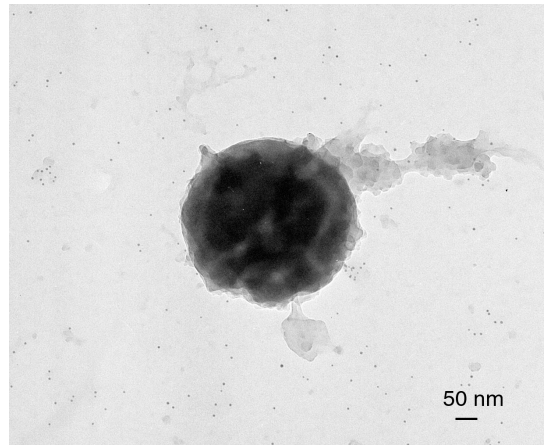
The minimization of (6.2) with respect to the unknown parameters  $\alpha^i$  and  $\beta^i$  and  $\mathbf{x}_j$  can be made with standard optimization tools. Of the tested three algorithms the classical Gauss–Newton algorithm was the most efficient while the Levenberg–Marquardt (Levenberg 1944, Marquardt 1963) and trust region subspace method (Byrd, Schnabel and Schultz 1988) performed relatively slowly. In practice, one must ignore too short chains to keep the memory use and computation time tolerable. In our experiments we did not consider chains shorter than 15.

## 7.5 Experiments

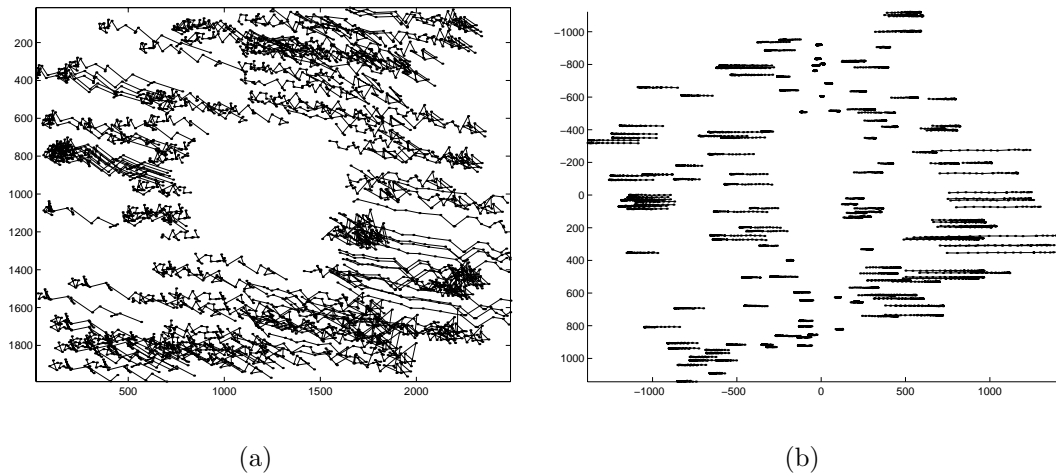
The first tested tilt series represented a mitochondrion where the size of the images was about  $2500 \times 2000$  pixels. The reference image is shown in Fig. 7.5. From the image series, 152 marker chains longer than 14 were found while 10 chains covered the whole image series. The standard deviation of the residual was 0.93 pixels, which is proof of the success of the algorithm. Note that deviation is not only caused by the uncertainty in marker coordinates but also deformation of the

---

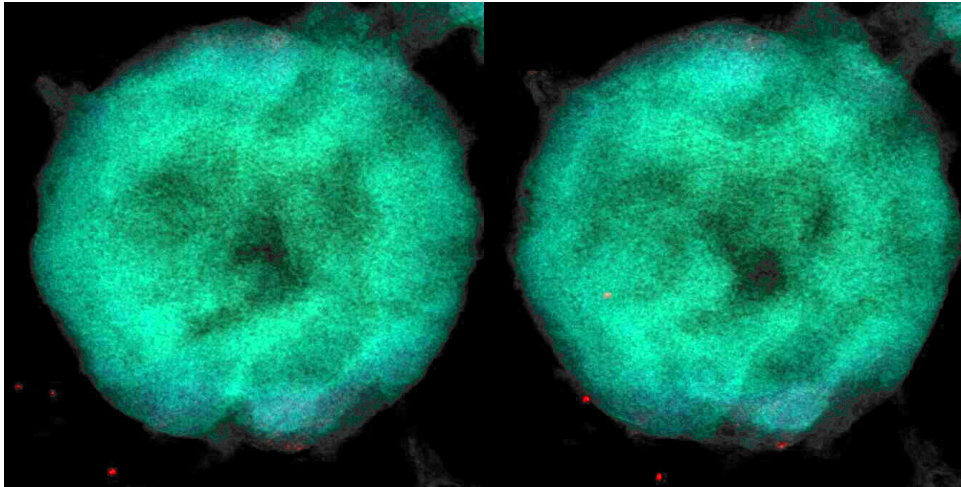
<sup>3</sup>Some marker coordinates may be taken multiple times because if the chain is broken in some image, the marker subchains are considered different. This does not, however, affect the validity of the approach because, as said, the coordinate origin may be chosen arbitrarily.



**Figure 7.5:** One image of the tilt series representing a whole-mounted critical-point-dried (cpd) mitochondrion, prepared as in Engelhardt (2000). The images have been taken at  $50\,000\times$  magnification with 120 kV voltage. The average diameter of a gold bead is 5–10 nm which corresponds to 10–20 pixels in the image.



**Figure 7.6:** Found gold marker trajectories after the transformation of the images to a common coordinate plane before (a) and after (b) the optimization. One chain corresponds to one marker location in several images. Ideally, the trajectories should be horizontal lines which is well satisfied here. Only the chains longer than 15 are plotted here for the sake of clarity.



**Figure 7.7:** Stereo image pair of the reconstructed mitochondrion.

specimen support during the imaging and inaccuracy of the goniometer<sup>4</sup>. To visualize the results, we have plotted some trajectories of the longest marker chains after their coordinates are transformed to a common coordinate plan (Fig. 7.6b). For comparison, the initial trajectories are shown in Fig. 7.6a. Ideally the markers should move along the epipolar planes, i.e., the trajectories would ideally be horizontal. The stereo image pair of the reconstructed 3D object is shown in Fig. 7.7. The reconstruction is made via the well-known maximum entropy method (MEM) (Lawrence, Jaffer and Sewell 1978).

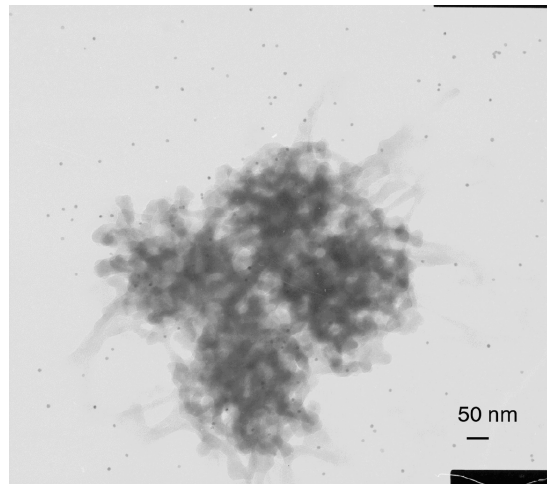
The second test image series represented a chromosome shown in Fig. 7.8. Here, 9 marker chains covered the whole image series and the minimum chain length was exceeded by 97 chains. The registration resulted in a standard deviation of 1.3 pixels in the residual, which again is an indication of success. The gold marker trajectories are plotted in Fig. 7.9, and the stereo image pair of the reconstructed object is shown in Fig. 7.10.

## 7.6 Further Development

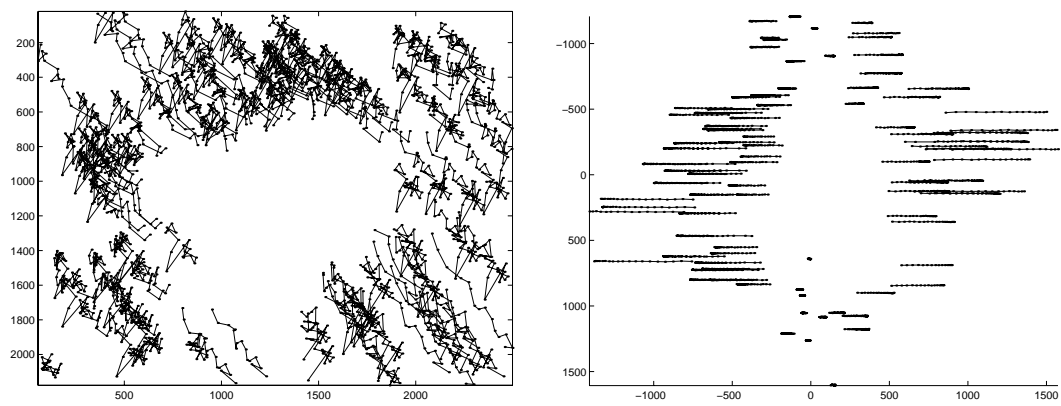
The purpose of this section is to give some ideas about how the proposed automatic alignment method can be developed even further. We discuss completing the marker chains by the rejected markers and combining separated subchains, handling the possible deformation of the sample, and dual-axis tilting.

---

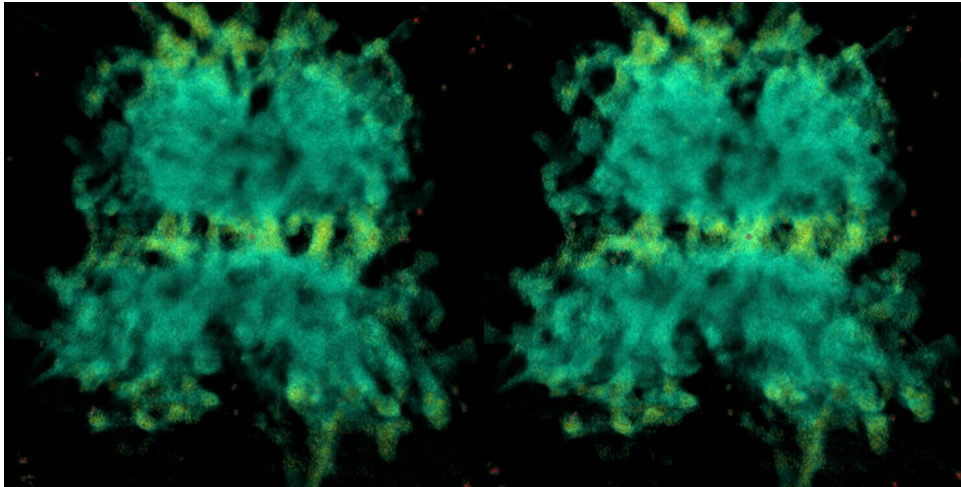
<sup>4</sup>In practice, the tilting axis is not fixed but fluctuates randomly. This contradicts the assumption of the motion model.



**Figure 7.8:** Reference image of the second test image series representing a whole-mounted cpd chromosome, prepared as described in Engelhardt (2000). The images are acquired with  $48\,000\times$  magnification and 100 kV voltage. The average gold markers diameter is 10 nm which corresponds to 20 pixels in the image.



**Figure 7.9:** Gold marker trajectories for the second chromosome image series before (a) and after (b) the optimization. Also here the trajectories are close to horizontal which indicates good results. For clarity no longer trajectories than 15 are plotted.



**Figure 7.10:** Stereo image pair of the reconstructed chromosome. A corresponding manually aligned reconstruction can be found in Engelhardt (2000).

The graph matching stage rejects some markers in images due to the threshold in the neighborhood correlation value (Section 7.3.2), and they are thus not utilized in the marker chains. If there are relatively few marker chains found and a higher level of accuracy in alignment is needed, practically all the missing markers of the marker chains can be identified by introducing a second stage to the algorithm. In the first pass, we obtain, in addition to the motion parameters, the three-dimensional coordinates of the markers. Using the model (7.5), the optimized three-dimensional marker coordinates can be projected onto each image where the marker for the chain has not been identified. A new connection is found if the coordinates of the reprojection and some marker coincide on the image plane. Broken chains and subchains may consequently be completed and connected.

If there are lots of markers in the images and the specimen is not significantly deformed during the imaging, there is no need for the second stage described above. However, the second pass can give the medicine to cure some problems caused by the deformation of the imaged specimen. After all the possible marker chains are found and subchains have been connected, there is no theoretical obstacle for user interaction to specify the portion of the images that are to be considered. Then, only those markers that lie inside the selected volume should be used in optimizing the final transformation parameters. In optimization, the parameters obtained in the first phase are naturally a good initial guess. If the observed deformation is intended to be handled automatically, one could use some robust regression techniques (see Chapter 2) in the second phase optimization.

The proposed method can be extended to the dual-axis tomography<sup>5</sup> (Penczek, Marko, Buttle and Frank 1995, Mastronarde 1997, McEwen and Marko 1999) in a straightforward way. One needs only to treat the two axes separately in finding and tracking the gold markers in them. Thereafter the two images that are closest to each other in the two series (usually the reference images) should be selected. Between these images, one needs to compute the epipolar geometry and find, using the tracking approach, which marker chains are equivalent in the two tilt series. As far as the optimization stage is concerned, there is no difficulty in solving all the parameters at once.

## 7.7 Summary and Discussion

We have proposed a novel way to register electron tomography images. Compared to the previous automated approaches, the suggested method is able to achieve a high level of accuracy because gold beads are used as feature points and they can be localized precisely; on the other hand, no manual picking of the markers is required any longer. In contrast to the previous automatic methods, the proposed method is an approach from the computer vision viewpoint, incorporating some recent advances in vision research.

Some more detailed advantages of the proposed method are, e.g., the fact that gold beads are found with high confidence, allowing the use of a huge number of markers, say hundreds in an image, in solving the alignment. The methods that are based on intensity thresholding are not able to find as many markers in an image at all. The use of epipolar geometry with the proposed graph matching procedure makes the tracking procedure reliable as the probability of obtaining a false correspondence becomes negligible.

There are also no obstacles in using the proposed method in dual-axis tomography via a straightforward modification. We have also discussed the evident ways of incorporating the information of practically all the markers shown in the images. Even though it is not usually necessary to implement the proposed second phase, in cases of partial deformation of the specimen it can provide a significant advantage.

Being a sophisticated and accurate method for tomographic image registration, the price is paid in the computational cost of the technique. The registration of one tilt series on a common computer workstation may take hours. However, computational cost is far cheaper than the time of a researcher who should do the manual picking of the gold beads from the images.

A disadvantage of the use of gold markers is that it is not always possible to include them in the preparation. In addition, the gold beads interfere with the

---

<sup>5</sup>In dual-axis tomography, the specimen is tilted over two different axes instead of one.

three-dimensional reconstruction of the object. In the next chapter, we therefore study the automatic registration problem when there are no fiducial markers in the images.



## Chapter 8

# TEM Image Acquisition Without Markers<sup>1</sup>

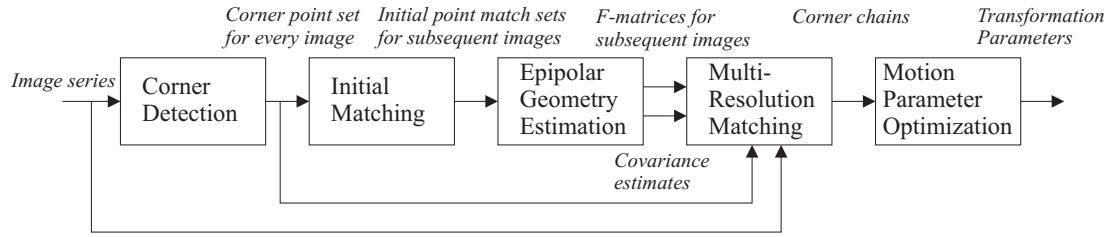
In the previous chapter, we obtained accurate results in automatic transmission electron microscope image alignment by using colloidal gold beads as fiducial markers. In general, if the use of markers has not been possible for some reason, the only option has been automatic cross-correlation-based registration methods. However, these methods are inaccurate and, as we will show, inappropriate for the problem as a whole. Conversely, in this chapter we propose a novel method that uses the 3D motion model but works without fiducial markers in the images. The method is based on matching and tracking the corner points of the intensity surface by first solving the underlying geometrical constraint of consecutive images in the tilt series. The results show that our method approaches the level of accuracy of the gold marker alignment and hence opens the way for new opportunities in the analysis of electron tomography reconstructions, particularly when markers cannot be used.

### 8.1 Introduction

As stated in the previous chapter, accurate alignment has been previously achieved by using conventional colloidal gold beads as fiducial markers since being spherical-shaped they can be well localized. Sometimes, however, it is not possible to include gold markers in the images. Moreover, gold markers interfere with computing the three-dimensional reconstruction. Therefore, it would occasionally be necessary to solve image alignment without markers. The electron tomogra-

---

<sup>1</sup>The preliminary steps to the markerless alignment were taken in Brandt and Heikkonen (2000*c*) but the actual conference and journal versions of this chapter are Brandt, Heikkonen and Engelhardt (2002) and Brandt, Heikkonen and Engelhardt (2001*b*), respectively.



**Figure 8.1:** Chart describing the proposed alignment method. Input and output values of each stage are displayed in italics.

phy literature reports automatic alignment using cross-correlation (Guckenberger 1982, Frank and McEwen 1992), where no fiducial markers are needed. However, that approach is based on 2D cross-correlation that is capable of modeling 2D plane motion only—not the 3D motion of the object. Since the implicit motion model of this approach is insufficient for the alignment problem as a whole, the result is not optimal at all and alignment errors accumulate along the image series.

In this chapter, we introduce a framework for registering TEM images which for the first time provides automatic alignment using the 3D motion model without any fiducial markers. The proposed approach is based on the techniques proposed earlier in this thesis, i.e., first estimating the epipolar geometries of the consecutive image pairs (Chapter 3), after which found feature points are matched and tracked using the wavelet-based multi-resolution approach (Chapter 5). The maximum likelihood estimates for the transformation parameters are finally obtained by solving the non-linear optimization problem described in the previous chapter. This chapter thus completes the thesis by showing how the proposed algorithms can efficiently solve the alignment problem where, above all, neither user-interaction nor fiducial markers are needed.

## 8.2 Methods

The principal intention of our method for automatic TEM image alignment is to track certain interest points in the image series and use these feature point tracks in estimating the 3D motion parameters of the object. However, the tracking is not a trivial problem: first, because the interest points should be those that can be accurately localized, second, the corresponding points should also be found in the other images, and third, the determined correspondences should not contain false matches. Our approach to the problem is summarized in Fig. 8.1.

Widely used feature points in tracking are corners that can be defined as the high-curvature points of the image intensity surface where the direction of the

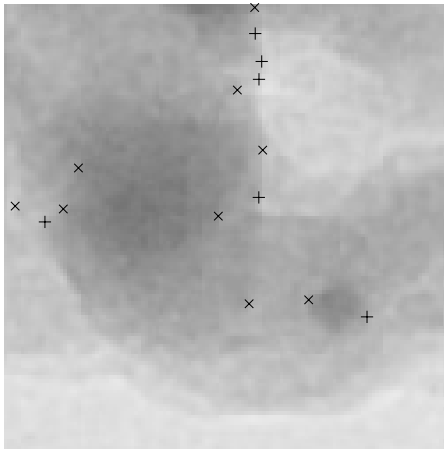
gradient changes rapidly. Numerous corner detectors exist in the literature (e.g. Moravec 1979, Kitchen and Rosenfeld 1982, Zuniga and Haralick 1983, Heitger, Rosenthaler, von der Heydt and Peterhans 1992) but the most frequently considered is perhaps the Harris detector (Harris and Stephens 1988) because of its good repeatability and localization accuracy (Schmid et al. 2000). We therefore used the improved Harris detector (Schmid et al. 2000) here and match and track the points as proposed in Chapter 5.

At this point, the reader might speculate as to whether such features as corners are appropriate feature points for electron microscope images, or whether they exist at all in images of small-scale biological objects. Although the use of the term “corner” is relatively stable in the computer vision community, it is a little misleading. A better term would be “interest point” since the corner points we obtain, for instance, with the Harris detector do not have to be such physical corners as we would normally consider. The term “interest points” fits better because the information content of the intensity neighborhood of Harris corner points is considerable larger than the neighborhood of a random point in the image (Schmid et al. 2000). Figure 8.2 shows an example of the corner features found in electron microscope images.

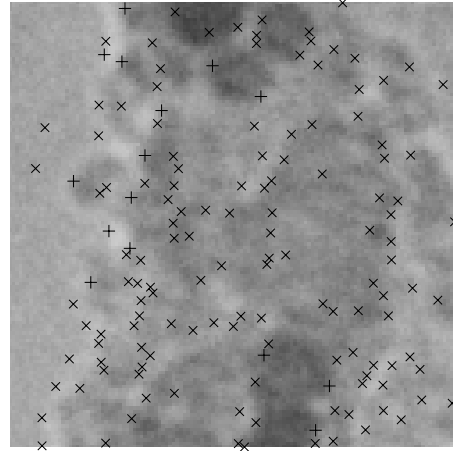
The wavelet-based multi-resolution matching technique (Chapter 5) forms the key point of our method. As our experiments showed, the probability of obtaining mismatches is practically very small. If more reliability is required, we could additionally consider three view correspondences, estimate the affine trifocal tensor, and reject the mismatches on the basis of the point transfer error (Torr 1995, Hartley and Zisserman 2000). Nevertheless, in this work we did not deem this necessary. After the correspondences between the interest point sets have been identified, it is straightforward to track the feature points through the image series. The found chains are thereafter used in the transformation parameter optimization. The optimization problem is solved as proposed in Chapter 7.

## 8.3 Experiments

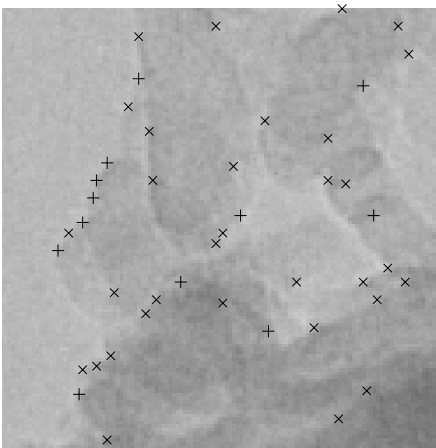
In this section, we demonstrate how the proposed method performs in practice. We will analyze four different tilt series. The first series is the same whole-mounted cpd chromosome that was used in the previous chapter. The gold marker aligned images provide us a reference to which we may compare our new method though we do not use the markers. The second series is an example from a situation where the markers are sprinkled on the preparation but in practice it is not even possible to perform a manual alignment. To demonstrate the viability of the proposed method we have reconstructed these sets and in addition two other tilt series that are absolutely free of markers. Specimen preparations are performed as described in Engelhardt (2000). As the reconstruction method we



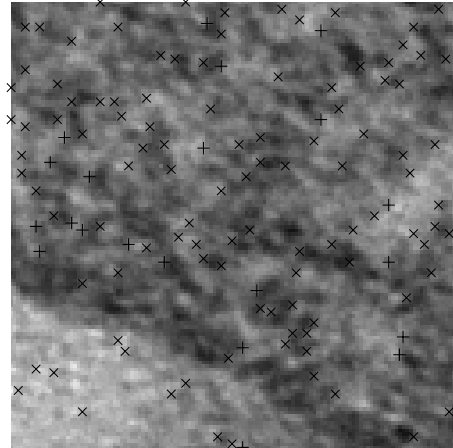
(a) Chromosome scaffold



(b) Microvillus



(c) Mitochondrion



(d) Section of the mitochondrion

**Figure 8.2:** Examples of Harris corners found in the four reference images used in the experiments, see Fig. 8.3. The “x” indicates a found corner and “+” a corner for which a matching point has been found in next image. The size of these detail windows is  $125 \times 125$  pixels. As can be seen, in the image of the mitochondrion section the noise level is very large. In any of the images, there are not really such sharp corners as one might expect; therefore the term interest point rather than corner might apply better. On the whole, these examples show that reliable corner point matching is far from a trivial problem.

use the maximum entropy method. (Lawrence et al. 1978), as in the previous chapter.

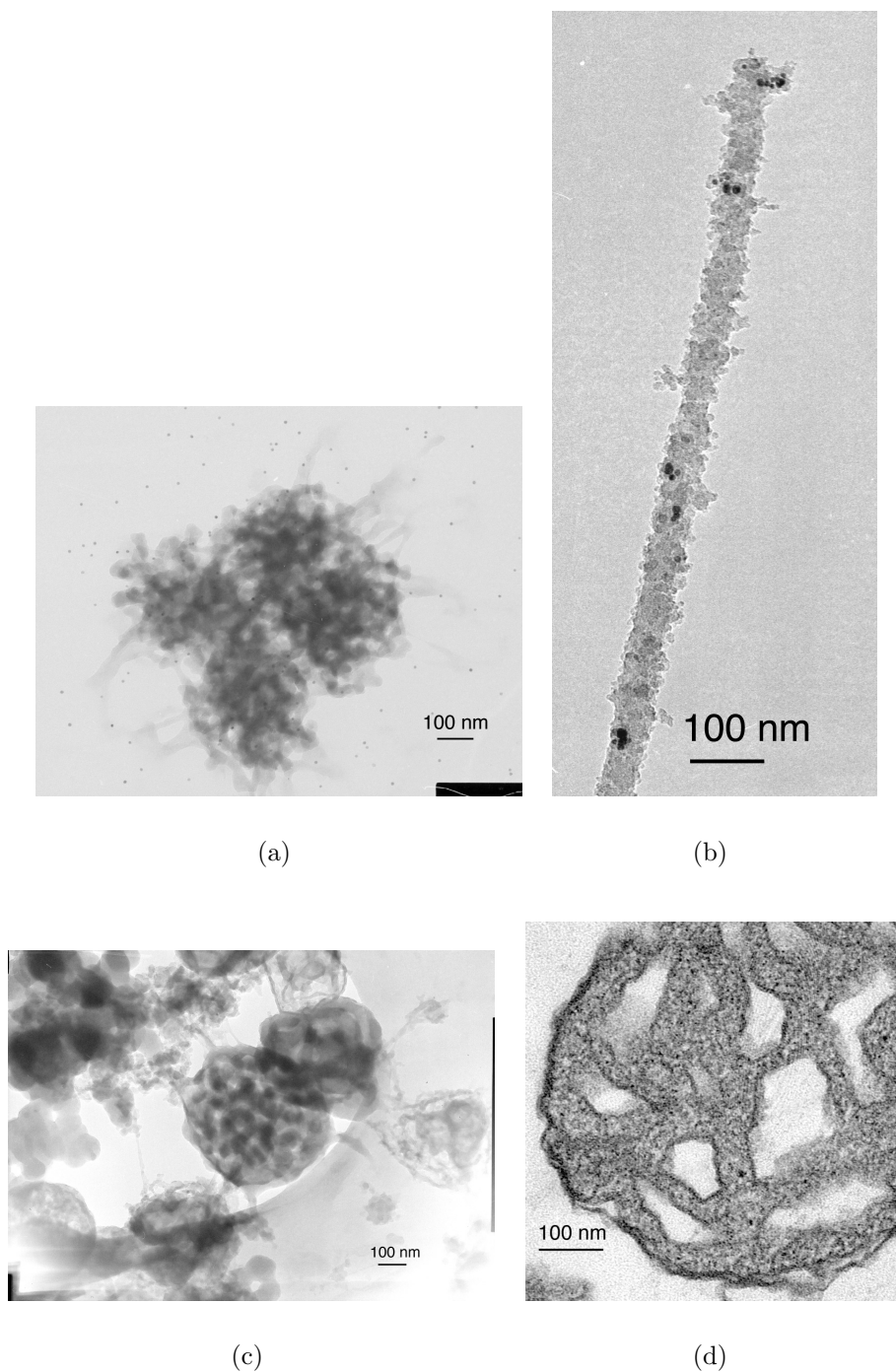
To evaluate the alignment we approximate the RMS (root mean square) estimation error that gives the average distance between the estimated coordinates and true (noise free) values, i.e., the mean error per data point. Assuming i.i.d. Gaussian noise with deviation  $\sigma$  in the image coordinates and using a linear approximation of the model, with  $N_i$  independent parameters, around the true observation vector in the observation space  $N_t$ , where  $N_t$  is the total number of independent measurements, the RMS estimation error is obtained from

$$\epsilon_{\text{est}} = \sigma \sqrt{\frac{N_i}{N_t}}, \quad (8.1)$$

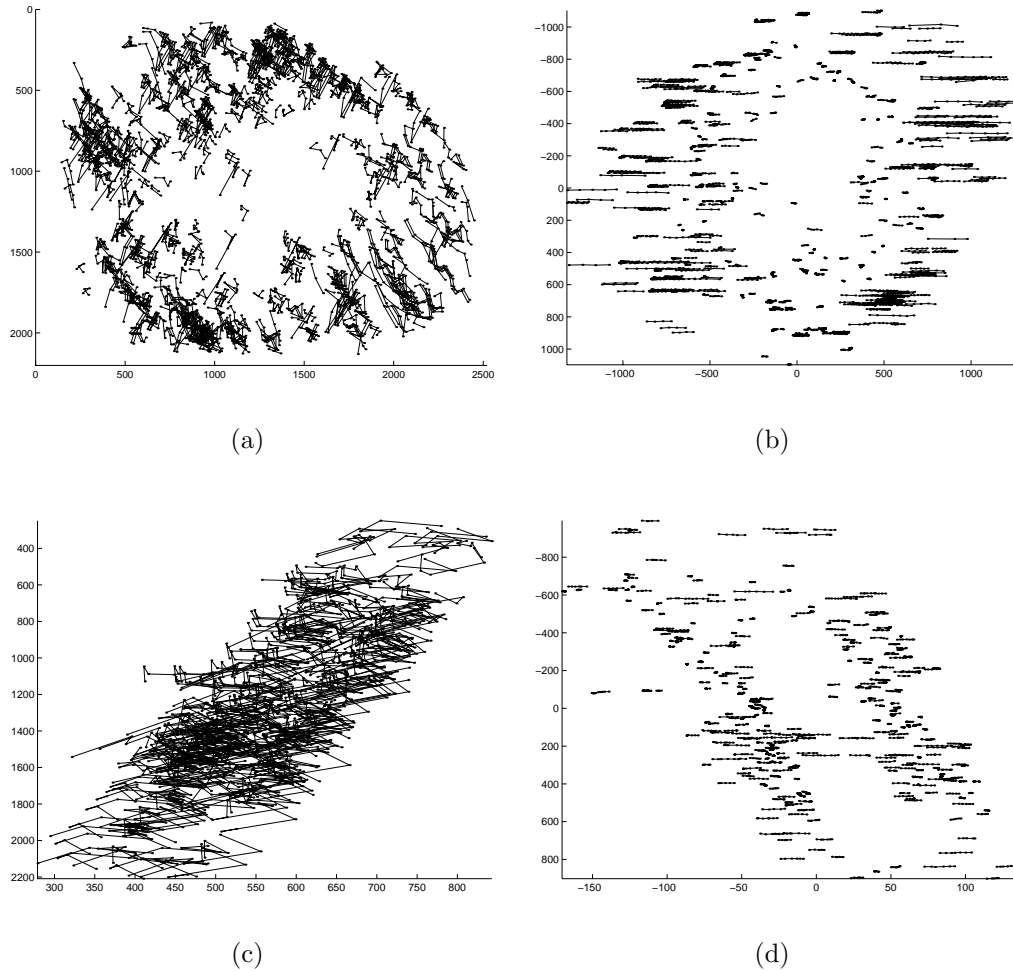
where we also approximate  $\sigma$  by its sample estimate. In our case,  $N_i = 5N + 3M - 5$ , where  $N$  is the number of images,  $M$  is the number of corner or marker tracks found, and  $N_t = 2 \sum_{i,j} \delta_{ij}$ . Since the RMS error approximates the average *true* error while it takes the dimensionality of the parameter space and the number of measurements into account, smaller RMS error should also imply more accurate alignment, independent of the alignment method or image content. It is therefore used here for evaluation between different alignment settings and tilt series.

Let us first analyze the alignment of the whole-mounted cpd chromosome scaffold tilt series (Fig. 8.3a). Technical details concerning the tilt series are collected in Table 8.1. By using gold markers, we obtained 1.3 pixels for the reprojection error, i.e., for the deviation between the estimated noise-free coordinates and the measured noisy counterparts. For the RMS estimation error we obtained 0.39 pixels. The alignment without the use of the gold markers resulted in  $\sigma = 2.5$  pixels for reprojection error; hence, it indicates that the noise in the coordinates was about twice as much. The RMS estimation error gave 1.4 pixels, which is a reasonable result too. It is not quite as good as was obtained by gold markers since the markers can be much more accurately localized and many of them cover the whole sequence; hence the number of parameters to be solved was much less in the marker case. The reconstruction of the whole chromosome scaffold is shown in Fig. 8.5a to allow comparison with not only the corresponding automatically aligned version (Fig. 7.10) but also the manually aligned reconstruction (Engelhardt 2000, Fig. 9). The corner trajectories are plotted in Figs. 8.4a and 8.4b (cf. Fig. 7.9).

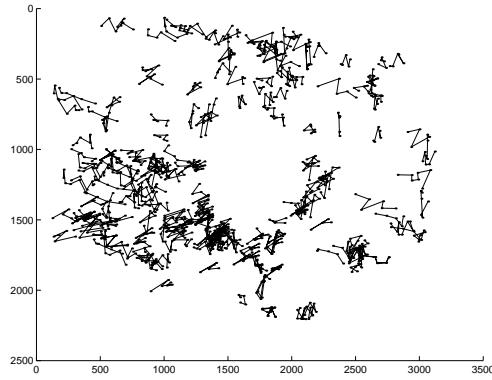
The second test set is a typical example where the proposed alignment method could be applied. The object in Fig. 8.3b represent a freely dangling microvillus of a whole-mounted cpd cell that is intended to be aligned with gold markers. However, the number of separated markers is too small and nearly all markers have been clustered, due to immunolabeling, such that even a manual picking tended to be practically impossible. The proposed method was therefore applied to the set and as a result the corner point deviation was found to be 0.78 pixels



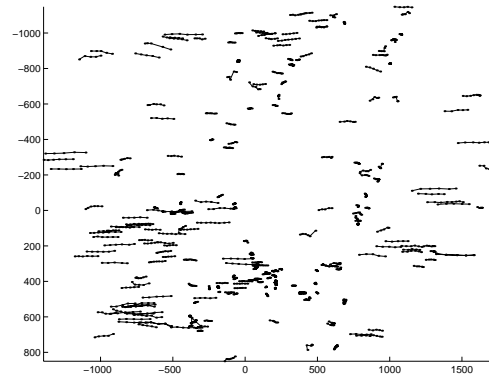
**Figure 8.3:** Reference images of the four image series used in the experiments. (a) Chromosome scaffold (aligned using markers in Chapter 7); (b) freely dangling microvillus with immunologically gold labeled mucin (3 nm) and ezrin (6 nm) (cf. Engelhardt 2000); (c) whole-mounted mitochondrion spanning a hole of the supporting film of the grid; and (d) section of an epoxy resin-embedded condensed mitochondrion.



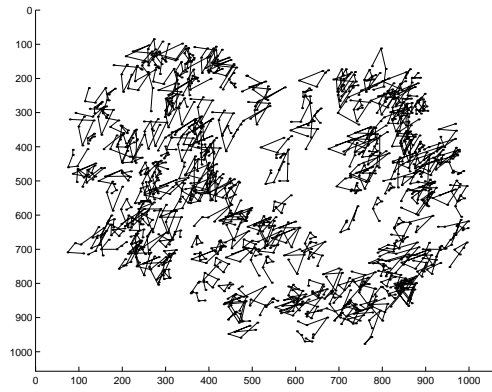
**Figure 8.4:** Corner trajectories for the second chromosome image series before (left column) and after (right column) the optimization. (a and b) Chromosome scaffold. (c and d) Microvillus. (e and f) Mitochondrion. (g and h) Section of the Mitochondrion. Trajectories are close to horizontal, indicating good results. As can be seen the tracks are relatively short. Median length was 4 or 5 in all cases though the shortest chains with length 2 and 3 were not used.



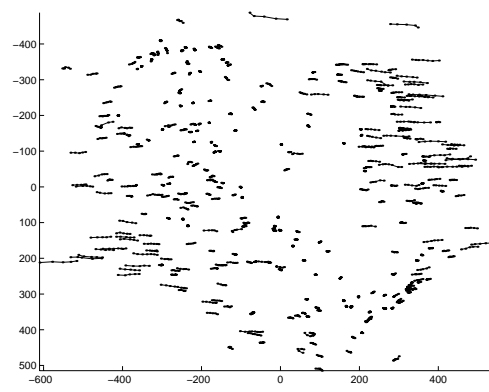
(e)



(f)



(g)



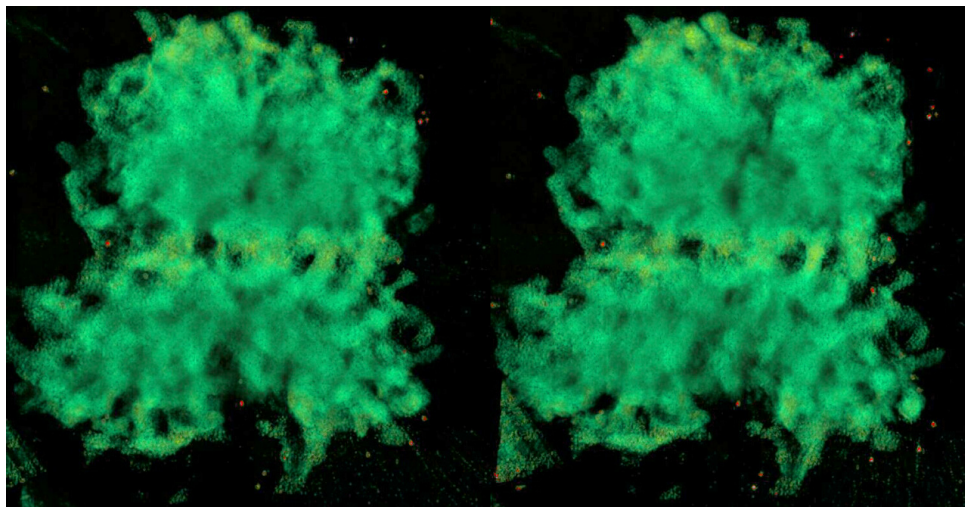
(h)

after the optimization. The corresponding RMS error was 0.45 pixels, which is a very good result considering the difficulty of the image; the region containing the information for the alignment in the images is very narrow. The corner trajectories and the corresponding reconstruction of the top part are shown in Figs. 8.4c and 8.4d and in Fig. 8.5b, respectively.

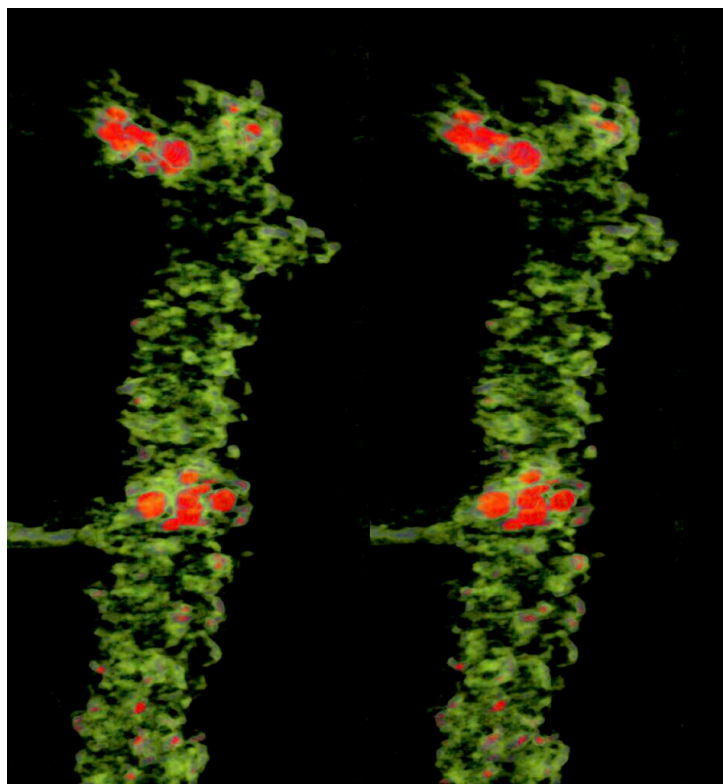
In Fig. 8.3c we have a whole-mounted cpd mitochondrion without markers. The alignment resulted in an estimate of  $\sigma_c = 3.4$  pixels for the noise deviation while the RMS error was 2.0 pixels. The larger deviation is due mostly to some serious dynamic deformation of the object<sup>2</sup> that can also be clearly seen from the tilt series. The trajectories also show that some tracks (Fig. 8.4f) are not perfectly straight after the optimization. However, the stereo image pair of the reconstructed mitochondrion in Fig. 8.5c shows that the proposed alignment method

<sup>2</sup>An object is here defined as a specimen with the specimen support.



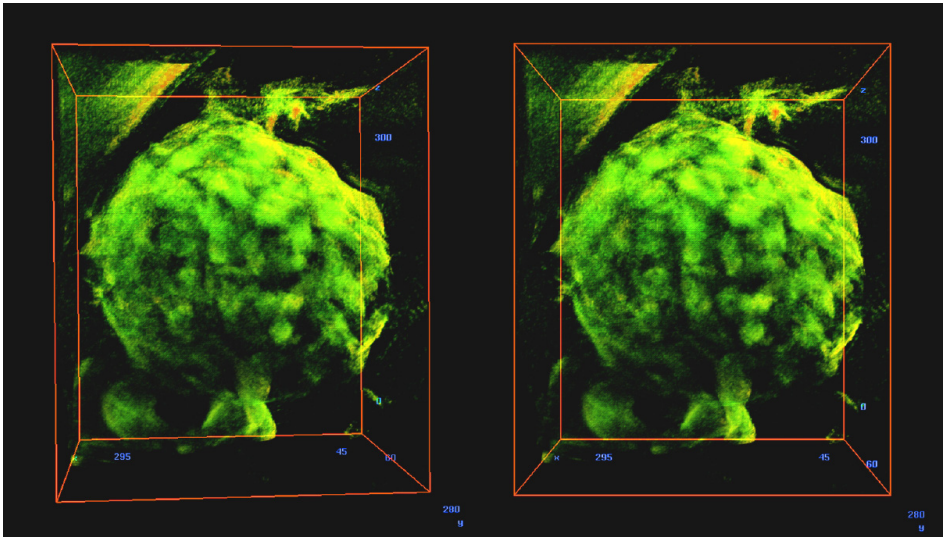


(a)

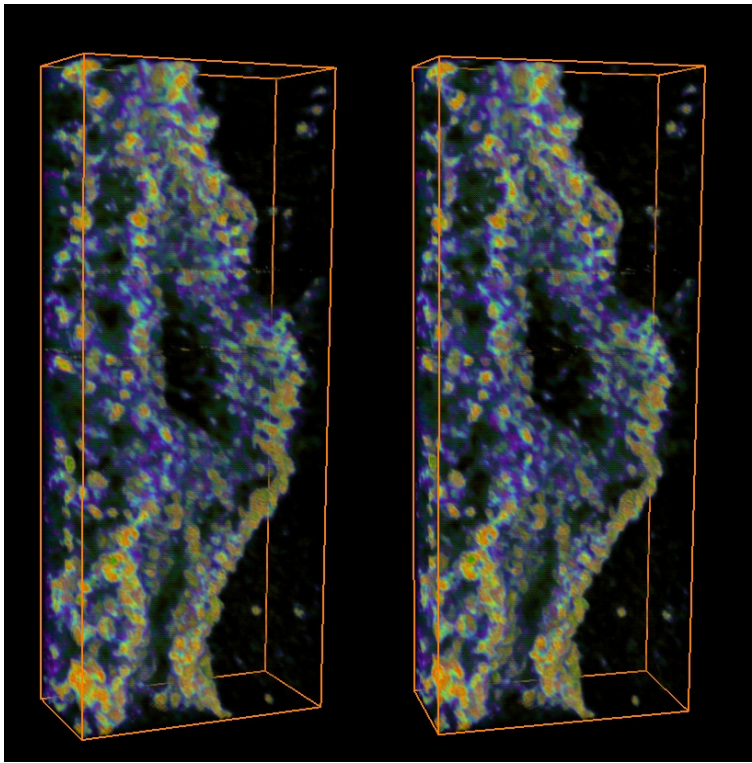


(b)

**Figure 8.5:** Reconstructions. (a) Chromosome scaffold (cf. Fig. 7.10); (b) reconstructed top part of the microvillus; (c) whole-mounted mitochondrion; and (d) reconstructed details of the outer region in the section of the mitochondrion.



(c)



(d)

**Table 8.1:** Technical details and results for the tilt series used in the experiments.

	Chr. (M)	Chr. (C)	Microvillus	Mitoch.	Section
Voltage (kV)	100	100	120	120	80
Magnification	48000	48000	50000	50000	40000
Image size	$2500 \times 2200$	$2500 \times 2200$	$1200 \times 2300$	$3200 \times 2300$	$1100 \times 1100$
Pixel width ( $\text{\AA}$ )	4.4	4.4	5.1	5.1	6.4
No. images ( $N$ )	41	41	41	40	36
No. chains ( $M$ )	109	627	796	248	324
No. measurements ( $N_t$ )	5602	6882	7716	2774	3554
No. parameters	527	2081	2588	939	1147
No. corners per image	-	1100	4300	7700	9300
median chain length	26	5	4	5	5
min chain length	15	4	4	5	5
max chain length	41	20	14	12	9
Reprojection err. ( $\hat{\sigma}$ )	1.3	2.5	0.78	3.4	1.3
RMS error ( $\hat{\epsilon}_{\text{est}}$ )	0.39	1.4	0.45	2.0	0.71

M = aligned with markers

C = aligned with corners

has worked out successfully. For better accuracy, the dynamic behavior of the object might be taken care of by robust regression techniques, as discussed in the following section.

The last tilt series is of a slice of another mitochondrion shown in Fig. 8.3d. Some images in the series were, however, so out of focus that only 36 images were finally used. Now, the alignment resulted in 1.3 pixels for the noise deviation estimate. The RMS error was 0.71 pixels. The result is surprisingly accurate even though the noise level in the images is considerably large, as can be seen from the reference image (see Fig. 8.2d). The trajectories in Figs. 8.4g and 8.4h and the stereo image pair of the reconstructed part in Fig. 8.5d verify that the alignment has been successful. The results are summarized in Table 8.1.

## 8.4 Summary and Discussion

We have proposed a novel way of aligning transmission electron microscope images automatically where no fiducial markers are needed. The solution lies in first solving the underlying epipolar constraint between consecutive images after which corner features are matched and tracked by the wavelet-based multi-resolution approach. The method utilizes the apparent motion model in the tilting operation in contrast to the previous cross-correlation-based methods with the implicit 2D plane motion model that are invalid for describing the three-dimensional motion.

Under Section 8.3 (Experiments), four tilt series were aligned using the proposed

approach where the series of chromosome scaffolds had also been automatically aligned with gold markers in the previous chapter. As the figure of merit we used the RMS estimation error that approximates the average error between the true and the estimated parameters and therefore allows direct comparison among the five individual alignment cases. The results suggest that under good imaging conditions the expected average error by the proposed technique is of the degree of a fraction of a pixel. In the case where there was also the gold marker alignment reference, better results were, however, obtained by using markers. For evaluation, the corner point trajectories were also plotted. They were reasonably close to horizontal after the optimization, which validates the applicability of the approach.

On the basis of the experiments we might draw the conclusion that it is recommended to use gold markers whenever possible. This is natural because gold beads are sphere-shaped and they can hence be localized very accurately. Moreover, the marker tracks are typically very long, implying that the relative number of parameters to be solved is smaller. Corner points can be localized quite accurately, but the tracks are typically much shorter than obtained with gold beads. Each track involves three additional parameters to be estimated and therefore even if we had the same number of measurements with same noise level, the marker alignment would be better in the expected level of accuracy.

The method proposed in this chapter is worth using especially in cases where it is not possible to use gold markers for alignment for some reason. In any case, the level of accuracy does not differ much when the proposed method is compared to the previous automatic method with markers. Hence, the level of accuracy achieved here is promising and future developments might even obviate the use of markers fully. In addition, small alignment errors are less important in practice if the final resolution of the reconstruction is to be decimated down by some factors, as oversampling is a commonly used technique in electron tomography.

The results obtained here for the corner deviation are additionally a bit pessimistic for some unnecessary technical reasons. As pointed out by McEwen and Marko (1999), it is recommended to start the tilting from one extreme tilt end and to move by constant increments into the other. Our data were not collected in this way, but started from the central position  $0^\circ$  and tilted to  $\pm 60^\circ$ . Since the goniometer is not an ideal apparatus, the motion model of a fixed tilt axis fails especially around the zero angle and a clear “bump” can be seen in the image series. The corner tracking is more vulnerable to this inaccuracy because of the shorter chain lengths, whereas the long marker chains have an obvious regularizing effect on the motion estimates such that the whole sequence is better aligned in total.

There are still many ways of improving the alignment results in the future. By increasing the number of measurements the expected deviation from true motion

parameters decreases. By estimating the affine trifocal tensor for image triplets one could establish more point matches without any retardation of confidence by using the point transfer (Hartley and Zisserman 2000). The affine trifocal tensor could be also used in pinpointing mismatches as well as connecting corner chains that have been broken. More point matches can also be obtained by not only matching consecutive image pairs but also every second image, etc. Line or curve segment matches provide also constraint over several views (Schmid and Zisserman 2000, Kahl and Heyden 1999).

For a better level of accuracy, robust motion estimation is an important issue that should be deliberated in solving the motion. Partial deformation of the imaged object during the imaging violates the assumption of i.i.d. Gaussian noise. The noise model is also inadequate if the established correspondences consist of mismatches. These problems could be overcome by replacing the square cost in (6.2) by a proper robust cost function that also tolerates large residual values. By introducing an unconstrained motion model instead of (7.5) problems caused by unstable tilt-axis could be solved.

To obtain even better reconstructions one might yet consider the following techniques. The proposed method could be expanded to double-tilt tomography (Penczek et al. 1995, Mastronarde 1997, McEwen and Marko 1999) similarly as proposed in Chapter 7. Conical tilting instead of axial tilting would also be attractive since the image sequence could be circularly matched and the accumulated error caused by short corner chains should also be smaller. It is also straightforward to combine the use of markers and corner points if, for example, there are some markers in the images but not enough. In the optimization one should then just minimize the Mahalanobis distance to take the unequal noise deviation between corners and markers into account.

To summarize, we have proposed a framework of aligning electron microscope images that for the first time uses the true 3D motion model but does not need any fiducial markers. Above all, it is the inevitable way for aligning the images in the most accurate way if there are no markers available in the images. Since the present implementation is just at the prototype level, there are many obvious ways to develop the method for better reliability, accuracy, and possible real-time computational efficiency. Since the level of accuracy obtained so far is close to what has been achieved by using fiducial markers, the use of markers should become unnecessary from the alignment point of view when the discussed ways of development are followed. This issue is left for future research.



## Chapter 9

# Conclusions

From the scientific point of view, perhaps the most important result of this thesis is the proposed maximum likelihood robust estimator. It is notable that an estimator can be consistent to the true parameter values even when there are outliers in the data and only a parametric model for their residual distribution is needed. As the key observation is that there is no need for the ill-posed hard decision for a residual to be correct or false from the parameter estimation point of view, the proposed estimator reflects the common sense principle (Cherkassky and Mulier 1998): *Do not attempt to solve a specified problem by indirectly solving a harder general problem as an intermediate step*. When the parametric form of the outlier residual distribution is unknown, the robust regression problem is a problem of statistical modeling rather than statistical estimation since the complexity of the residual model must be determined to deal with a finite number of observations.

In the projective F-matrix estimation, the model selection verified the known fact that even simple models, with two kernel mixtures, may produce good results in small-sample estimation even though the “true” distribution was far more complex. Consequently, the number of observations may be too small to determine the fine-structure of the outlier distribution. As most accurate robust estimation is an important goal in itself, another notable by-product of the F-matrix estimation is that the covariance matrix of the F-matrix can be computed without having to classify the residual observations. Even though it seems that the finite uncertainty of F-matrix estimates has not significantly been taken into consideration so far, perhaps partly because of this earlier difficulty in its estimation, we see that, in general, well based uncertainty considerations are inevitable for computer vision itself if it is to establish its status as a science.

Since the whole problem relating to whether the covariance matrix of the fundamental matrix contains disparity information of the scene has been controversial, the understanding of the linear approximation of the epipolar line covariance can

be seen as an important result. As far as the matching of two images is concerned, the derived most probable point is the point where one should have its matching window centered. Moreover, we believe that the derived probability density for point–point correspondence from the fundamental matrix covariance is one of the most important results of the thesis. Since the probability density takes both epipolar geometry and its uncertainty into account, it is a useful tool for matching, and it could be used, for instance, as a prior distribution in computing the dense disparity field of the scene along with the photometric constraints in the future.

Our methods for affine reconstruction with missing data may be useful for multi-frame tracking applications since the Tomasi–Kanade factorization algorithm works only without missing data. Though our iterative algorithms give the ML solution with missing data when the starting point is sufficiently close to the maximum, a challenging problem for the future remains its solving in closed-form or, alternatively, showing why the closed-form solution does not exist. Nevertheless, our intermediate solution for structure and translations, given the inhomogeneous projection matrices, is useful in bundle adjustment at least in computing the initial guess for the iterative search. In addition, the proposed method of degeneracy identification may find applications.

From the electron tomography application point of view, this work has been fruitful. With the proposed methods, TEM image alignment with markers can be performed with a high level of accuracy since practically all the fiducial markers can be utilized. The second algorithm for those cases where no markers can be used is built around the proposed multi-resolution matching technique, and it should give better results than the previous primitive approaches in the ET field. It is also encouraging that, as outlined, there are still many possibilities for future development of the methods for improved efficiency and a better level of accuracy. As more accurate alignment implies better reconstructions, the entire electron tomography field could benefit from the results of this thesis, if the methods are distributed for practical use.

To summarize, the aims set for this work have been achieved and new scientific knowledge has been revealed without stepping too far outside the original framework, even though the research problem may be considered quite classical in the motion estimation sense. This is very encouraging since it implies that there are many more unsolved problems to occupy future research in the geometric branch of computer vision.



# Bibliography

- Akaike, H. (1977), On entropy maximization principle, *in* P. R. Krishnaiah, ed., ‘Applications of Statistics’, North Holland, Amsterdam, pp. 27–41.
- Alvarez, L., Deriche, R., Sánchez, J. and Weickert, J. (2000), Dense disparity map estimation respecting image discontinuities: A pde and scale-space based approach, *in* ‘Proceedings of the IAPR Workshop on Machine Vision Applications’, Tokyo, Japan, pp. 423–426.
- Beran, R. (1974), ‘Asymptotically efficient adaptive rank estimates in location models’, *Ann. Stat.* **2**(1), 63–74.
- Biernacki, C., Celeux, G. and Govaert, G. (2000), ‘Strategies for getting the largest likelihood in mixture models’, Presented at the invited session in Joint Statistical Meetings 2000, Indianapolis. Presentation available at <http://www.inrialpes.fr/is2/people/celeux/>.
- Bishop, C. M. (1998), *Neural Networks for Pattern Recognition*, Oxford University Press.
- Boufama, B. and Mohr, R. (1995), Epipole and fundamental matrix estimation using virtual parallax, *in* ‘Proceedings of the 5th International Conference on Computer Vision’, pp. 1030–1036.
- Brandt, S. (1999), Use of shape features in content-based image retrieval, Master’s thesis, Helsinki University of Technology.
- Brandt, S. (2002a), Closed-form solutions for affine reconstruction, *in* ‘Proceedings of the Statistical Methods in Video Processing Workshop, in conjunction with ECCV 2002’, Copenhagen, Denmark, pp. 109–114.
- Brandt, S. (2002b), Maximum likelihood robust regression with known and unknown residual models, *in* ‘Proceedings of the Statistical Methods in Video Processing Workshop, in conjunction with ECCV 2002’, Copenhagen, Denmark, pp. 97–102.

- Brandt, S. and Heikkonen, J. (2000a), Automatic alignment of electron tomography images using markers, *in* D. P. Casasent, ed., 'Intelligent Robots and Computer Vision XIX: Algorithms, Techniques, and Active Vision', Vol. 4197 of *Proceedings of SPIE*, Boston, MA, pp. 277–287.
- Brandt, S. and Heikkonen, J. (2000b), 'A Bayesian weighting principle for the fundamental matrix estimation', *Pattern Recognit. Lett.* **21**(12), 1081–1092.
- Brandt, S. and Heikkonen, J. (2000c), A fully automatic alignment of electron tomography images without fiducial markers, *in* 'Proceedings of IAPR Workshop on Machine Vision Applications', Tokyo, Japan, pp. 307–310.
- Brandt, S. and Heikkonen, J. (2000d), A new robust Bayesian method for the affine F-matrix estimation, *in* B. Girod, G. Greiner, H. Niemann and H.-P. Seidel, eds, 'Vision, Modeling, and Visualization 2000', Saarbrücken, Germany, pp. 39–46.
- Brandt, S. and Heikkonen, J. (2001a), Multi-resolution matching of uncalibrated images utilizing epipolar geometry and its uncertainty, *in* 'Proceedings of the IEEE International Conference on Image Processing', Vol. 2, Thessaloniki, Greece, pp. 213–216.
- Brandt, S. and Heikkonen, J. (2001b), Optimal method for the affine F-matrix and its uncertainty estimation in the sense of both noise and outliers, *in* 'Proceedings of the Eighth International Conference on Computer Vision', Vol. 2, Vancouver, Canada, pp. 166–173.
- Brandt, S., Heikkonen, J. and Engelhardt, P. (2001a), Automatic alignment of transmission electron microscope images using markers, *in* 'Proceedings of the 12th Scandinavian Conference on Image Analysis', Bergen, Norway, pp. 33–40.
- Brandt, S., Heikkonen, J. and Engelhardt, P. (2001b), 'Automatic alignment of transmission electron microscope tilt series without fiducial markers', *J. Struct. Biol.* **136**(3), 201–213.
- Brandt, S., Heikkonen, J. and Engelhardt, P. (2001c), 'Multiphase method for automatic alignment of transmission electron microscope images using markers', *J. Struct. Biol.* **133**(1), 10–22.
- Brandt, S., Heikkonen, J. and Engelhardt, P. (2002), On the alignment of transmission electron microscope images without fiducial markers, *in* 'Proceedings of the 16th International Conference on Pattern Recognition', Quebec City, Canada, pp. 278–281.
- Burrus, C. S., Gopinath, R. A. and Guo, H. (1998), *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall.

- Byrd, R., Schnabel, R. and Schultz, G. (1988), ‘Approximate solution of the methods for large scale nonlinear minimization over two-dimensional subspaces’, *Math. Program.* **40**, 247–263.
- Canny, J. (1986), ‘A computational approach to edge detection’, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698.
- Celeux, G. and Diebolt, J. (1985), ‘The sem algorithm: A probabilistic teacher algorithm derived from the em algorithm for the mixture problem’, *Computat. Stat. Quart.* **2**, 73–82.
- Celeux, G. and Govaert, G. (1992), ‘A classification em algorithm for clustering and two stochastic versions’, *Computat. Stat. Data Anal.* **14**, 315–332.
- Cherkassky, V. and Mulier, F. (1998), *Learning From Data*, John Wiley & Sons.
- Chojnacki, W., Brooks, M., van den Hengel, A. and Gawley, D. (2000), ‘On the fitting of surfaces to data with covariances’, *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1294–1303.
- Chojnacki, W., Brooks, M., van den Hengel, A. and Gawley, D. (2002), A new approach to constrained parameter estimation applicable to some computer vision problems, in D. Suter, ed., ‘Proceedings of the Statistical Methods in Video Processing Workshop, in conjunction with ECCV 2002’, pp. 43–48.
- Cohen, J. and Roth, M. (1976), ‘On the implementation of Strassen’s fast multiplication algorithm’, *Acta Inform.* **6**, 341–355.
- Cormen, T. H., Leiserson, C. E. and Rivest, R. L. (1996), *Introduction to Algorithms*, 17th edn, The MIT Press, Cambridge, MA.
- Csurka, G., Zeller, C., Zhang, Z. and Faugeras, O. D. (1997), ‘Characterizing the uncertainty of the fundamental matrix’, *Comput. Vis. Image Underst.* **68**(1), 18–36.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), ‘Maximum likelihood from incomplete data via the EM algorithm (with discussion)’, *J. R. Statist. Soc. Ser. B-Stat. Methodol.* **39**, 1–38.
- Engelhardt, P. (2000), Electron tomography of chromosome structure, in R. A. Meyers, ed., ‘Encyclopedia of Analytical Chemistry’, Vol. 6, John Wiley & Sons Ltd, pp. 4948–4984.
- Faugeras, O. (1993), *Three Dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press.
- Faugeras, O. and Luong, Q.-T. (2001), *Geometry of Multiple Images*, The MIT Press.

- Fischler, M. and Bolles, L. (1981), ‘Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography’, *Commun. ACM* **24**, 381–385.
- Fraley, C. and Raftery, A. E. (1998), ‘How many clusters? which clustering method? answers via model-based cluster analysis’, *Comput. J.* **41**, 578–588.
- Frank, J., ed. (1992), *Electron Tomography; Three Dimensional Imaging with the Transmission Electron Microscope*, Plenum Press.
- Frank, J. and McEwen, B. F. (1992), Alignment by cross-correlation, in J. Frank, ed., ‘Electron Tomography; Three Dimensional Imaging with the Transmission Electron Microscope’, Vol. 8, Plenum Press, New York, USA, pp. 205–213.
- Fung, J. C., Liu, W., Ruijter, W. D., Chen, H., Abbey, C. K., Sedat, J. W. and Agard, D. A. (1996), ‘Toward fully automated high-resolution electron tomography’, *J. Struct. Biol.* **116**(1), 181–189.
- Golub, G. H. and Loan, C. F. V. (1989), *Matrix Computations*, second edn, The John Hopkins University Press, chapter 7, pp. 345–346.
- Golub, G. H. and Loan, C. F. V. (1996), *Matrix Computations*, third edn, The John Hopkins University Press.
- Gonzalez, R. C. and Woods, R. E. (1993), *Digital Image Processing*, Addison Wesley.
- Guckenberger, R. (1982), ‘Determination of a common origin in the micrographs of tilt series in three-dimensional electron microscopy’, *Ultramicroscopy* **9**, 167–174.
- Han, K. A., Lee, J. C. and Hwang, C. J. (1994), Image clustering using self-organizing feature map with refinement, in ‘Proceedings of the IEEE International Conference on Neural Networks’, Vol. 1, pp. 465–469.
- Hannah, M. J. (1989), ‘A system for digital stereo image matching’, *Photogramm. Eng. Remote Sens.* **55**(12), 1765–1770.
- Harris, C. and Stephens, M. (1988), A combined corner and edge detector, in ‘4th Alvey Vision Conference’, pp. 147–151.
- Hartley, R. I. (1993), Euclidean reconstruction from uncalibrated views, in ‘Proceedings of the Second Europe–U.S. Workshop on Invariance’, Ponta Delgada, Azores, pp. 187–202.

- Hartley, R. I. (1994), ‘Projective reconstruction and invariants from multiple images’, *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(10), 1036–1041.
- Hartley, R. I. (1997), ‘In defense of the eight-point algorithm’, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(6), 580–593.
- Hartley, R. and Strum, P. (1997), ‘Triangulation’, *Comput. Vis. Image Underst.* **68**(2), 146–157.
- Hartley, R. and Zisserman, A. (2000), *Multiple View Geometry in Computer Vision*, Cambridge University Press.
- Heitger, F., Rosenthaler, L., von der Heydt, R. and Peterhans, E. (1992), ‘Simulation of neural contour mechanisms: from simple to end-stopped cells’, *Vision Res.* **32**(5), 963–981.
- Hoff, W. and Ahuja, N. (1989), ‘Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection’, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(2), 121–136.
- Hogg, R. V. (1974), ‘Adaptive robust procedures: A partial review and some suggestions for future applications and theory’, *J. Am. Stat. Assoc.* **69**(348), 909–927.
- Huber, P. J. (1981), *Robust Statistics*, Wiley.
- Huynh, D. and Heyden, A. (2001), Outlier detection in video sequences under affine projection, in ‘Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, Vol. 1, pp. 695–701.
- INRIA Syntim Project (1999), <URL: <http://www-syntim.inria.fr/syntim/analyse/paires-eng.html>>.
- Irani, M. and Anandan, P. (2000), Factorization with uncertainty, in ‘Proceedings of the Sixth European Conference on Computer Vision’, Dublin, Ireland, pp. 539–553.
- Jacobs, D. (1997), Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images, in ‘Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, pp. 206–212.
- Jacobs, D. W. (2001), ‘Linear fitting with missing data for structure-from-motion’, *Comput. Vis. Image Underst.* **82**, 57–81.
- Kahl, F. and Heyden, A. (1998), Using conic correspondences in two images to estimate the epipolar geometry, in ‘Proceedings of the 6th International Conference on Computer Vision’, Mumbai, India, pp. 761–766.

- Kahl, F. and Heyden, A. (1999), ‘Affine structure and motion from point, lines, and conics’, *Int. J. Comput. Vis.* **33**(3), 163–180.
- Kanatani, K. (1996), *Statistical Optimization for Geometric Computation*, Vol. 18 of *Machine Intelligence and Pattern Recognition*, Elsevier Science, Amsterdam.
- Kitchen, L. and Rosenfeld, A. (1982), ‘Gray-level corner detection’, *Pattern Recognit. Lett.* **1**, 95–102.
- Lawrence, M. C. (1992), Least-squares method of alignment, in J. Frank, ed., ‘Electron Tomography; Three Dimensional Imaging with the Transmission Electron Microscope’, Plenum Press, New York, USA, chapter 8, pp. 197–204.
- Lawrence, M., Jaffer, M. and Sewell, B. (1978), ‘The application of the maximum entropy method to electron microscopic tomography’, *Ultramicroscopy* **31**, 285–302.
- Levenberg, K. (1944), ‘A method for the solution of certain problems in least squares’, *Quart. Appl. Math.* **2**, 164–168.
- Lew, M. S. and Huang, T. S. (1999), Optimal multi-scale matching, in ‘Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, pp. 88–93.
- Liu, H. and Bhattacharya, P. (2000), Uncalibrated stereo matching using DWT, in ‘Proceedings of the 15th International Conference on Pattern Recognition’, Vol. 1, Barcelona, Spain, pp. 114–118.
- Longuet-Higgins, H. (1981), ‘A computer algorithm for reconstructing a scene from two projections’, *Nature* **293**, 133–135.
- Luong, Q.-T., Deriche, R., Faugeras, O. and Papadopoulos, T. (1993), On determining the fundamental matrix: Analysis of different methods and experimental results, Technical report, INRIA, Sophia-Antipolis, France.
- Luong, Q.-T. and Faugeras, O. (1996), ‘The fundamental matrix: Theory, algorithms, and stability analysis’, *Int. J. Comput. Vis.* **17**(1), 43–76.
- Ma, S. (1993), ‘Conic-based stereo, motion estimation, and pose determination’, *Int. J. Comput. Vis.* **1**, 7–25.
- Marquardt, D. (1963), ‘An algorithm for least-squares estimation of nonlinear parameters’, *SIAM J. Appl. Math.* **11**, 431–441.
- Marr, D. and Poggio, T. (1979), ‘A computational theory of human stereo vision’, *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* pp. 301–328.

- Mastronarde, D. (1997), 'Dual-axis tomography: an approach with alignment methods that preserve resolution', *J. Struct. Biol.* **120**, 343–352.
- McEwen, B. F. and Marko, M. (1999), 'Three-dimensional transmission electron microscopy and its application to mitosis research', *Methods Cell Biol.* **61**, 81–111.
- Moravec, H. (1979), Visual mapping by a robot rover, in 'Proceedings of the 5th International Joint Conference on Artificial Intelligence', pp. 598–600.
- Morris, D. D. and Kanade, T. (1998), A unified factorization algorithm for points, line segments and planes with uncertainty models, in 'Proceedings of the 6th International Conference on Computer Vision', Bombay, India, pp. 696–702.
- Penczek, P., Marko, M., Buttle, K. and Frank, J. (1995), 'Double-tilt electron tomography', *Ultramicroscopy* **60**, 393–410.
- Poelman, C. J. and Kanade, T. (1997), 'A paraperspective factorization method for shape and motion recovery', *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(9), 206–218.
- Ponce, J. and Genc, Y. (1996), Epipolar geometry and linear subspace methods: A new approach to weak calibration, in 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 776–781.
- Pratt, W. K. (1991), *Digital Image Processing*, John Wiley & Sons, Inc., New York.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1993), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press.
- Quan, L. and Kanade, T. (1996), A factorization method for affine structure from line correspondences, in 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 803–808.
- Reid, I. and Murray, D. (1996), 'Active tracking of foveated feature clusters using affine structure', *Int. J. Comput. Vis.* **18**(1), 41–60.
- Ress, D., Harlow, M. L., Schwarz, M., Marshall, R. M. and McMahan, U. J. (1999), 'Automatic acquisition of fiducial markers and alignment of images in tilt series for electron tomography', *J. Electron Microsc.* **48**(3), 277–287.
- Rissanen, J. (1978), 'Modeling by shortest data description', *Automatica* **14**, 465–471.

- Rissanen, J. (1983), 'A universal prior for integers and estimation by minimum description length', *Ann. Stat.* **11**(2), 416–431.
- Rissanen, J. (1989), *Stochastic Complexity in Statistical Inquiry*, Vol. 15 of *Series in Computer Science*, World Scientific, Singapore.
- Rousseeuw, P. J. (1984), 'Least median of squares regression', *J. Am. Stat. Assoc.* **79**, 871–880.
- Rousseeuw, P. J. and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, Wiley.
- Sacks, J. (1975), 'An asymptotically efficient sequence of estimators of a location parameter', *Ann. Stat.* **3**(2), 285–298.
- Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986), *Akaike Information Criterion Statistics*, KTK Scientific Publishers.
- Schmid, C., Mohr, R. and Bauckhage, C. (2000), 'Evaluation of interest point detectors', *Int. J. Comput. Vis.* **37**(2), 151–172.
- Schmid, C. and Zisserman, A. (2000), 'The geometry and matching of curves over multiple views', *Int. J. Comput. Vis.* **40**(3), 199–233.
- Schwarz, G. (1978), 'Estimating the dimension of a model', *Ann. Stat.* **6**(2), 461–464.
- Shapiro, L. and Brady, J. (1995), 'Rejecting outliers and estimating errors in an orthogonal regression framework', *Philos. Trans. R. Soc. Lond. Ser. A-Math. Phys. Eng. Sci.* **350**.
- Shapiro, L., Zisserman, A. and Brady, M. (1994), Motion from point matches using affine epipolar geometry, in 'Proceedings of the Third European Conference on Computer Vision'.
- Shum, H.-Y., Ikeuchi, K. and Reddy, R. (1995), 'Principle component analysis with missing data and its application to polyhedral object modeling', *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(9), 854–867.
- Srinath, M., Rajasekaran, P. and Viswanathan, R. (1996), *Introduction to Statistical Signal Processing with Applications*, Prentice Hall, chapter 5, pp. 146–149.
- Stone, C. J. (1975), 'Adaptive maximum likelihood estimators', *Ann. Stat.* **3**(2), 267–284.
- Tomasi, C. and Kanade, T. (1992), 'Shape and motion from image streams under orthography: A factorization approach', *Int. J. Comput. Vis.* **9**(2), 137–154.



- Torr, P. (1995), Motion Segmentation and Outlier Detection, PhD thesis, University of Oxford.
- Torr, P. and Murray, D. (1993), Outlier detection and motion segmentation, *in* ‘Sensor Fusion VI’, Proceedings of SPIE, Boston, MA, pp. 432–443.
- Torr, P. and Murray, D. (1997), ‘The development and comparison of robust methods for estimating the fundamental matrix’, *Int. J. Comput. Vis.* **24**(3), 271–300.
- Torr, P. and Zisserman, A. (1998), Concerning Bayesian motion segmentation, model averaging, matching and the trifocal tensor, *in* ‘Proceedings of the 5th European Conference on Computer Vision’, Freiburg, Germany, pp. 511–527.
- Torr, P. and Zisserman, A. (2000), ‘MLESAC: A new robust estimator with application to estimating image geometry’, *Comput. Vis. Image Underst.* **78**(1), 138–156.
- Triggs, B. (2001), Joint feature distributions for image correspondence, *in* ‘Proceedings of the 8th International Conference on Computer Vision’, Vol. II, Vancouver, Canada, pp. 201–208.
- Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A. (2000), Bundle adjustment – a modern synthesis, *in* B. Triggs, A. Zisserman and R. Szeliski, eds, ‘Vision Algorithms: Theory and Practice’, Vol. 1883 of *LNCS*, Springer-Verlag, Berlin/New York, pp. 298–372.
- Ullman, S. (1979), *The Interpretation of Visual Motion*, The MIT Press, Cambridge, MA.
- Vapnik, V. N. (2000), *The Nature of Statistical Learning Theory*, second edn, Springer-Verlag, New York.
- Wei, G.-Q., Brauer, W. and Hirzinger, G. (1998), ‘Intensity- and gradient-based stereo matching using hierarchical gaussian basis functions’, *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1143–1160.
- Weng, J., Huang, T. and Ahuja, N. (1992), ‘Motion and structure from line correspondences: Closed-form solution, uniqueness, and optimization’, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(3), 318–336.
- Weng, J., Huang, T. S. and Ahuja, N. (1989), ‘Motion and structure from two perspective views: Algorithms, error analysis, and error estimation’, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(5), 451–476.
- Wiberg, T. (1976), Computation of principal components when data are missing, *in* ‘Proceedings of the Second Symposium on Computational Statistics’, pp. 229–236.

- Wu, C. J. (1983), ‘On the convergence properties of the EM algorithm’, *Ann. Stat.* **11**(1), 95–103.
- Xu, G. and Zhang, Z. (1996), *Epipolar Geometry in Stereo, Motion and Object Recognition*, Vol. 6 of *Computational Imaging and Vision*, Kluwer Academic Publishers.
- Zhang, Z. (1998*a*), ‘Determining the epipolar geometry and its uncertainty: A review’, *Int. J. Comput. Vis.* **27**(2), 161–195.
- Zhang, Z. (1998*b*), ‘On the optimization criteria used in two-view motion analysis’, *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(7), 717–729.
- Zhang, Z., Deriche, R., Faugeras, O. and Luong, Q. (1994), ‘A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry’, *Artif. Intell.* **78**, 87–119.
- Zhang, Z. and Loop, C. (2001), ‘Estimating the fundamental matrix by transforming image points in projective space’, *Comput. Vis. Image Underst.* **82**, 174–180.
- Zhuang, X., Wang, T. and Zhang, P. (1992), ‘A highly robust estimator through partially likelihood function modeling and its application in computer vision’, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(1), 19–35.
- Zuniga, O. A. and Haralick, R. M. (1983), Corner detection using the facet model, *in* ‘Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition’, pp. 30–37.



ISBN 951-22-6136-7 (printed)  
ISBN 951-22-6137-5 (PDF)  
ISSN 1455-0474